REPORT DOCUMENTATION PAGE	Form Approved OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
May 95

3. REPORT TYPE AND DATES COVERED

4. TITLE AND SUBTITLE
New Methods and Comparative Evaluations For Robust and Biased-Robust Regression Estimation

5. FUNDING NUMBERS

6. AUTHOR(S)
James Robert Simpson

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
AFIT Students Attending:
Arizona State University

8. PERFORMING ORGANIZATION REPORT NUMBER
AFIT/CI/CIA
95-014D

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
DEPARTNEMT OF THE AIR FORCE
AFIT/CI
2950 P STREET, BDLG 125
WRIGHT-PATTERSON AFB OH 45433-7765

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for Public Release IAW AFR 190-1
Distribution Unlimited
BRIAN D. GAUTHIER, MSgt, USAF
Chief Administration

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

14. SUBJECT TERMS

15. NUMBER OF PAGES
284

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT

18. SECURITY CLASSIFICATION OF THIS PAGE

19. SECURITY CLASSIFICATION OF ABSTRACT

20. LIMITATION OF ABSTRACT

NEW METHODS AND COMPARATIVE EVALUATIONS FOR ROBUST AND BIASED-ROBUST REGRESSION ESTIMATION

by

James Robert Simpson

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

ARIZONA STATE UNIVERSITY

May 1995

NEW METHODS AND COMPARATIVE EVALUATIONS FOR ROBUST AND

BIASED-ROBUST REGRESSION ESTIMATION

by

James Robert Simpson

has been approved

May 1995

APPROVED:

______________________________,Chairperson

____

J Bert Keats

____

Supervisory Committee

ACCEPTED:

Norma Hubele

____

Department Chairperson

Bianca Bernstein

____

Dean, Graduate College

ABSTRACT

Least squares estimation is the predominant technique for regression analysis due to its universal acceptance, elegant statistical properties, and computational simplicity. Unfortunately, the statistical properties that make least squares so powerful depend on several assumptions that are often violated using real data. The normally distributed errors assumption, which enables tests of regressor significance, is invalid if only a single outlying observation occurs in the data. Robust regression methods are less sensitive to outliers than the method of least squares. Recently published techniques suggest improved robust estimation performance. These robust methods are comparatively evaluated using Monte Carlo simulation. Evaluation results lead to new proposals from a class of robust methods called GM-estimators. GM-estimation constrains the excess influence that observations outlying in the regressor space have on parameter estimates, enabling fits to the majority of the data regardless of outlier location. Several GM-estimation proposals are developed and evaluated. Two preferred GM proposals are compared with top performing existing robust methods in a comprehensive study of outlier and nonoutlier configurations. The best performing methods are an existing technique called MM-estimation and a proposed GM technique. Both techniques perform well against a variety of dataset configurations.

Least squares estimation can also be adversely impacted by dependencies among the regressors called multicollinearity. The resulting least squares parameter estimates can change significantly with only slight changes in the data. Alternative techniques, called biased estimation methods, induce a small amount of bias in the estimates, resulting in large reductions in parameter estimate variability. The combined outlier-multicollinearity problem occurs frequently in routine data. Methods that address this problem effectively combine biased and robust estimation techniques. The robust GM proposal from this paper is used to develop a biased-robust method. Two previously published approaches are compared with the proposal in simulation experiments. The best performing published technique is also compared with the proposal using a dataset containing a cloud of outliers and severe multicollinearity. The proposed biased-robust method outperforms the published technique both in the simulation and the example.

ACKNOWLEDGEMENTS

I wish to thank all who have provided me support and assistance during my research. I am deeply grateful to have had the opportunity to work with my committee chairman, Dr. Douglas Montgomery. He has provided the inspiration, guidance, and instruction that I value greatly. I have benefited enormously from his continued support and confidence in my abilities. I would also like to thank Dr. J. Bert Keats for his equally valuable, unwavering support of my research. From him I have learned the importance of determination and perseverance in academic endeavors. I want to thank Dr. Ronald Askin for his helpful suggestions and guidance on this paper.

I am indebted to Dr. Alfio Marazzi for his immeasurable assistance with my programming software. He has freely shared too much of his time to respond to my constant volley of electronic messages. I also wish to thank Dr. Douglas Hawkins for his time and for sharing his expertise in robust methods.

I am forever indebted to my brother, Dr. Wendell Simpson, for suggesting that such an undertaking is possible and for ensuring it becomes a reality through his continued inspiration, counsel, friendship and love.

Most importantly, I would like to thank my wife, Kiki, not only for her unending love and support, but for being my partner through it all.

TABLE OF CONTENTS

Page

Chapter Page

Chapter Page

Chapter Page

Chapter Page

Appendix

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

1.1 Introduction and Background of the Problem

Regression analysis is a statistical applications tool that is useful in nearly all areas of engineering and science that require fitting models to sets of data. Although there are several methods available for estimating model parameters, the least squares method is used most often because of its general acceptance, elegant statistical properties and ease of computation. Unfortunately, the mathematical elegance that makes least squares so popular depends on a number of fairly restrictive and often unrealistic assumptions. The assumption that makes least squares so attractive in terms of general model hypothesis testing and parameter significance testing is that the distribution of the errors is normal or Gaussian. This assumption can be violated if one or more sufficiently outlying observations are present in the data, resulting in less than reliable estimates of the model parameters. A second condition that potentially impacts the reliability of least squares estimates is multicollinearity, which is a near-linear dependency among the regressors. Multicollinearity can cause large variability in the estimates of the parameters, sometimes resulting in parameter estimates that differ from the true values by orders of magnitude or have the incorrect sign.

Outliers, which occur often in real data, appear for many reasons including typing or computation errors, interchanging of values, inadvertent observations from different populations, or transient effects. Outliers can also be due to observations selected from genuinely long-tailed distributions. Hampel et al. (1986, p. 25) summarized the results of numerous studies of the frequency of outliers in real data and conclude that altogether 1-10% outliers in routine data are more the rule than the exception. Outliers are found in the response variable (y-variable) or the regressor variables (x-variables). Regardless of their origin, a single, sufficiently outlying observation in a data set can render least squares estimation meaningless. Robust regression estimation methods are designed to be less sensitive than least squares to outliers, resulting in improved fits to the nonoutlying observations. Ronchetti (1987) points out that the goal of a robust selection procedure is to choose a model which fits the majority of the data, taking into account that the errors may not be normally distributed. A number of robust regression estimation techniques have been proposed, and some have been successfully used in practice, but none of the techniques have been unanimously accepted by practitioners or by members of the research community. All of the robust methods developed to this point have weaknesses under certain outlier scenarios, or are poor performers relative to least squares when outliers are not present. Some recent developments in a class of robust methods called GM-estimators by Simpson et al. (1992) and Coakley and Hettmansperger (1993) have prompted an interest in finding variations that may perform better than all the existing robust methods.

Another condition that presents problems for least squares estimation is multicollinearity. Many regression model-building processes involve finding variables that behave similarly to the response. It is also common for these descriptive variables to behave similarly relative to each other, often times capturing similar information regarding the response. In this case, the descriptive or regressor variables have a degree of dependency or multicollinearity. Multicollinearity can make it difficult to accurately estimate the model parameters under least squares. Least squares estimation requires that the matrix of regressor variables **X** be multiplied by itself, forming **X′X**, and then inverted, resulting in $(\mathbf{X'X})^{-1}$. Near-linear dependency among the regressors can result in an ill-conditioned **X′X,** meaning that the matrix inversion routines can be very inaccurate, resulting in considerable error in the least squares estimates of the model parameters. In general, multicollinearity tends to inflate the variance and absolute value of the least squares coefficients. In this case, the main fault with the least squares estimate is that it forces the estimator to be unbiased. Alternative estimation techniques that have been proposed induce a little bias by augmenting the regressor variable matrix, causing increased stability in the **X′X** matrix, and resulting in large reductions in the variance of the estimates. Biased estimation methods, such as ridge regression, can provide stable coefficient estimates with computational ease.

Outliers and multicollinearity occur simultaneously in real data as often as each problem occurs separately. Relative to the hundreds of references in biased-only and robust-only techniques, the research in biased-robust regression has been sparse. The advances in the combined area have been made only in the last two decades by Holland (1973), Pariente and Welsch (1977), Hogg (1979), Askin and Montgomery (1980), Montgomery and Askin (1981), Pfaffenberger and Dielman (1985), Lawrence and Marsh (1984), Walker and Birch (1985, 1988), and Walker (1987). Askin and Montgomery (1984) and Pfaffenberger and Dielman (1990) have followed up the development of their techniques by performing Monte Carlo simulation studies to compare various approaches. The most common approach to biased-robust estimation is augmented weighted least squares which allows a biased estimator and robust estimator to be combined into a single biased-robust estimator. Many of the existing robust estimators can be easily combined with biased estimators using the augmented-weighted least squares approach.

In fact, several of the recently created biased-only and robust-only estimators are excellent candidates for an improved combined estimator.

1.2 Statement of the Problem

Frequently, difficulties arise when practitioners try to apply least squares regression estimation to real world data containing outliers or dependent regressors. The traditional view that least squares is robust to deviations from its assumptions of normally distributed errors and uncorrelated regressors prevents users from applying other, more appropriate methods. In instances where model adequacy diagnostics reveal a poor least squares fit due to outliers and/or collinearity, the practitioner is often not able to properly fit a model because robust or biased-robust estimation techniques are not known or available. The increasing presence of observational data with correlated regressors and abundant outliers makes advances in the state of the art of robust and biased-robust estimation essential.

The rapid development of alternative robust estimators has resulted in the need to evaluate and test these methods to determine their performance capabilities. Opportunities also exist for the development of improved robust estimators that can potentially fit the majority of outlier and nonoutlier situations. A class of robust estimators called GM-estimators provides the framework for the development of such a robust technique. Although progress continues to be made in making algorithms for robust methods available, there is a growing demand from users to have the proper tools available to implement when least squares fails. Improved methods must be computationally practical and available to software developers.

A need also exists to develop and test alternative approaches for the combined problem so that the community of practitioners is aware of the potential to accurately estimate regression model terms under realistic data conditions. The most recent advancements in robust-only and biased-only estimation warrant development of improved robust, and combined biased-robust estimators.

1.3 Research Objective

The objective of this research is to investigate possible advances in the field of robust regression estimation through the development of an alternative GM-estimator. This technique, or the best performing robust method, will then be used to develop and evaluate a superior performing biased-robust regression estimator. The proper accomplishment of this objective first requires a needed comprehensive evaluation of existing competing robust methods. Information gained in the competing methods evaluation will aid a well-directed approach in the development of new robust estimators. The newly developed estimators will then be compared with each other and with existing methods in a subsequent evaluation. The inability to develop an improved performing robust estimator will not detract from the significance of the research. A comprehensive evaluation of existing techniques has not been accomplished and is of tremendous value to those working with regression analysis. The best overall robust techniques, new or existing, will then be incorporated into a combined biased-robust estimator.

1.4 Scope and Outline of the Research

The purposes of this dissertation are to evaluate, develop and test robust regression methods and biased-robust regression methods. Although the focus of this research is on the development of new methods, a number of other integral tasks are associated with method development that are equally as important. For instance, in order to develop a new, improved technique, a thorough evaluation of existing techniques must be performed. In order to evaluate these existing methods, an understanding of possible outlier or collinearity data configurations must be obtained. As part of this process, it is important that all of the possible "messy" data configurations that expose vulnerabilities in the various techniques be investigated so that fair and comprehensive technique evaluations are performed. This type of development process is adopted for both the robust methods and the biased-robust methods. The process can be outlined for both types of methods as:

1. Gain an understanding (through pilot studies) of the different types of outlier and collinearity conditions that result in method estimation differences.
2. Develop the datasets to be used to evaluate and compare existing methods and new methods.
3. Screen the large number of potential existing methods. Use theoretical knowledge, previously published method comparisons and this study's own pilot experiments to trim the list down to the most promising methods.
4. Use the knowledge gained from the screening study to determine the best direction for developing improved techniques. Continue to use pilot simulation studies to test concepts as they are developed.
5. Evaluate the new alternatives in an aggressive study of their performance abilities. Select the best performing new methods for comparison in further studies that also include the candidate existing methods.
6. Pool existing and new alternatives for a comprehensive evaluation of performance. Include all possible clean and messy data configurations to fully test the methods.

7. Evaluate each technique's performance and select the few best performing techniques for a final examination.
8. Develop an application or example using real world data or a previously published example. Modify the data if necessary to fully challenge the estimators.
9. Perform the estimation of the application/example and evaluate the results. Comment on the performance and make a final recommendation for the best performing method.

The dissertation chapters are structured so that the process described above is apparent and, in general, is conducted in the sequence outlined. Following a discussion of the related literature in Chapter 2, Chapter 3 discusses the initial outlier data configuration experiments and the screening of the potential robust techniques. Chapter 4 deals with the development of alternative GM-estimators. These methods are developed by modifying the integral components of this multi-stage technique. A performance evaluation is conducted on ten robust alternatives and the best performing methods are recommended for further evaluation. Chapter 5 involves a detailed description and rigorous evaluation of existing and alternative robust methods. Methods are evaluated in terms of error of estimation relative to the true model coefficients. The best performing alternative method is fully developed in Chapter 6. This chapter also includes an application of this technique using a modified real set of satellite cost data. In Chapter 7, a biased-robust estimator is developed and compared to two previously published methods. This chapter also provides an example for technique comparison. Chapter 8 provides a summary and detailed discussion of the dissertation conclusions. Areas for future research are also recommended.

Chapter 2

Review of Robust and Biased Estimation Methods

In general, the majority of the research on alternatives to least squares estimation in the presence of outliers and correlated regressors has addressed either the outlier issue or the collinearity issue, but seldom has addressed the combined problem. This review of the related literature will consist of a brief discussion of least squares estimation followed by descriptions of related robust estimation methods, biased estimation methods and biased-robust methods. Robust methods are covered in more detail than biased methods because a major portion of this dissertation focuses on comparisons of existing robust methods and the development of new robust techniques. Also, a thorough understanding of the biased-robust estimation problem is attainable only by first becoming familiar with the work in robust-only and biased-only estimation methods. The contributions to biased-robust estimation follow naturally from components of the separate research areas and will be discussed in detail regarding both the estimation approaches and the Monte Carlo simulation comparisons.

2.1 Least Squares Estimation

Regression modeling is used to develop a theoretical or statistical explanation of physical phenomena under study. It normally involves transforming real data with explanatory variables and a response variable into a linear mathematical equation or expression. This linear statistical model is generally expressed in the form

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \varepsilon \tag{2.1}$$

where y is a random variable called the response; $\beta_0, \beta_1, \ldots, \beta_k$ are constants whose values are not known but are estimated from the experimental or observational data; $x_0, x_1, \ldots, x_k$ are mathematical variables with controlled or observed values; and ε is a random variable representing the unexplained random variations in the response.

The linear model can also be expressed in matrix form as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{2.2}$$

where **y** is a n x 1 vector of responses. The **X** matrix includes an initial column of 1's for the intercept so that the number of columns is k+1=p. So, **X** is an n x p matrix of the levels of the regressor variables; β is a p x 1 vector of model coefficient estimates which includes the intercept; and ε is an n x 1 vector of errors.

The primary role of least squares estimation in the development of linear models is to estimate the model coefficients β in a logical manner. To do this, we require only that the vector ε of random errors has mean **0** and variance $\sigma^2\mathbf{I}$. These requirements also imply that the random vector y has mean $\mathbf{X}\beta$ and variance-covariance matrix $\sigma^2\mathbf{I}$.

To obtain the least squares estimators for β in the linear model, we first must rewrite the linear model in the form

$$\mathbf{y} = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{e} \tag{2.3}$$

where $\hat{\beta}$ are estimators of β and **e** is not a vector of unobservable random errors, but a vector of observable residuals.

The objective of the method of least squares is to choose the estimators in such a way that the sum of the squares of the residuals is minimized. This objective is expressed as

$$\min_{\beta} \sum_{i=1}^{n} e_i^2 = \min_{\beta} \sum_{i=1}^{n} \left(y_i - \mathbf{x}_i \hat{\beta}\right)^2 \tag{2.4}$$

This objective is solved by taking the partial derivatives with respect to $\hat{\beta}$ and setting the equations to zero. This system of equations, also known as the normal equations, is of the form

$$(\mathbf{X}'\mathbf{X})\hat{\beta} = \mathbf{X}'\mathbf{y} \tag{2.5}$$

To solve for $\hat{\beta}$, multiply each side of the equation by $(\mathbf{X}'\mathbf{X})^{-1}$ to obtain the least squares estimators

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \tag{2.6}$$

It can be shown that these least squares estimators $\hat{\beta}$ are unbiased, meaning E($\hat{\beta}$)=β. The Gauss-Markov theorem shows that, among the class of linear unbiased estimators for β, the estimator $\hat{\beta}$ is the best in the sense that the variances of $\hat{\beta}$ are minimized. For this reason the least squares estimates are the best linear unbiased estimators (referred to as BLUE).

When the model errors are normally distributed, the method of least squares estimation is attractive in the sense that the estimate of β has additional desirable statistical properties. If we assume that the errors are normally distributed, we can show that the least squares estimates are also the maximum likelihood estimates and that these estimates are the uniformly minimum variance unbiased estimators. Under conditions of nonnormal distributions, particularly heavy-tailed error distributions, least squares no longer has these desirable properties. In fact, it can be shown that the maximum likelihood estimator for the heavy-tailed Laplace distribution is a robust technique called the least absolute value (LAV) estimator. In general, robust methods seek to minimize some function of the residuals that is less sensitive to outliers. Insensitivity is obtained by replacing the least squares objective (2.4) with a function of the residuals that is less gradual, perhaps of the form

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \rho(y_i - \mathbf{x}_i^{'}\hat{\boldsymbol{\beta}})$$

The function ρ is a function of the residuals that is often related to the likelihood function for an appropriate error distribution. The class of estimates using this approach is *M*-estimation. A number of ρ functions have been proposed, and although several perform well in general, no variation is universally the best. The objective function is normally minimized by using an iterative technique to solve the resulting nonlinear system of normal equations of the form

$$\sum_{i=1}^{n} \psi(e_i / s)\mathbf{x}_i = \mathbf{0}$$

where ψ is the derivative of ρ and $\mathbf{x}_i$ is the row vector of explanatory variables of the i^{th} case. An estimate of scale s is needed to standardize the residuals because the solution to the normal equations is not equivariant with respect to a magnification of the *y*-axis. The most common approach for solving this system is iteratively reweighted least squares (IRLS), resulting in an estimator of the form

$$\hat{\boldsymbol{\beta}} = (\mathbf{X'WX})^{-1}\mathbf{X'Wy}$$

where $\mathbf{W} = diag(w_1, w_2, \ldots, w_i)$ and $w_i = \psi(e_i / s) / (e_i / s)$. This approach is widely used because any ordinary least squares package can be used to perform the estimation. *M*-estimation will be described in more detail in the next section along with several other robust methods.

2.2 Robust Estimation

The issue of robustness goes back to the beginnings of statistics, most notably in the study of measures of location. In fact, Rey (1983) notes that the Greek besiegers of antiquity switched from using the mean to a more robust measure, the median. Thorough accounts of the early work in robust statistics can be found in papers by Harter (1974-1976), Huber (1972), and Stigler (1973). It was not until recent decades though that robust estimation became a true research area. The awareness was created by people such as E. S. Pearson, G. E. P. Box and J. W. Tukey. Box (1953) actually coined the term *robustness* and Tukey (1960) demonstrated the drastic nonrobustness of the mean and presented robust alternatives. In the 1960s, papers by Huber (1964, 1965) and Hampel (1968) formed the basis for the theory of robust estimation and extended this theory to applications such as regression.

Since these pioneering papers on robust estimation in regression, many approaches have been presented but no single approach is either optimum or superior to the others in all aspects. The important criteria used in the field to determine the strengths and weaknesses of an estimator will be introduced prior to the discussion of each of the techniques. Although some criteria are more important than others for a particular type of dataset, the ideal estimator would have the positive characteristics of all the following criteria.

Efficiency: Expressed as a percentage, the degree to which the estimator performs like least squares in the presence of Gaussian or normally distributed errors. The term is computed as the mean squared error of the least squares fit divided by the mean squared error of the robust fit. Efficiencies near 90-95% are desirable.

Breakdown point: The breakdown point of an estimator is the amount of contamination allowed in the data (usually a percentage or fraction) before the estimate ceases to give useful information about the parameters.

A formal finite-sample definition of breakdown is given in Rousseeuw and Leroy (1987, p. 9). Using a sample of *n* data points such that

$$Z = \left\{\left(x_{11},\ldots,x_{1p},y_1\right),\ldots,\left(x_{n1},\ldots,x_{np},y_n\right)\right\}$$

and let T be a regression estimator. Applying T to a sample Z yields regression estimator $T(Z) = \hat{\beta}$. Consider all possible corrupted samples Z' obtained by replacing any m of the original data points by arbitrary values. The maximum bias that can be caused by this contamination is

$$\text{bias}(m; T, Z) = \sup_{Z'} \left\| T(Z') - T(Z) \right\|$$

where the supremum is over all possible Z'. If bias(m; T, Z) is infinite, then the M outliers can have an arbitrarily large effect on T, which may be expressed by declaring the estimator breaks down. Thus, the breakdown point of the estimator T at the sample Z is defined as

$$bp_n^* = \min\left\{\frac{m}{n}; \text{bias}(m; T, Z) \text{ is infinite}\right\}$$

Breakdown points can be as low as $1/n$ (or sometimes referred to as 0%) meaning that only a single outlying observation can cause an estimator to be meaningless, as is the case with least squares. Breakdown points can also be as high as $n/2$ (or 50%), meaning that up to half of the data can be contaminated and the estimator can still be useful.

Bounded Influence: A characteristic of a robust method referring to its ability to control the amount of impact that points outlying in the **X**-space have on model estimation. These outlying

X-space points are often called leverage or high leverage points, because they tend to "pull" the model fit in their direction. Least squares is most susceptible to high leverage points, but some robust methods also have unbounded influence.

Determining whether or not an estimator has bounded influence is obtained a study of the influence function. The influence function (Hampel 1974) describes the effect of an additional observation in any point $\mathbf{x}_i$ on a statistic *T*, given a large sample distribution *F*. The influence function IF(X; *T*, *F*) is the first derivative of the statistic *T* at an underlying distribution *F* (*T*(*F*)). An example of *T*(*F*) for *M*-estimators in regression is defined implicitly in:

$$\int \mathbf{X}'\psi\left(\mathbf{y}-\mathbf{X}T(F)\right)dF = 0$$

The corresponding influence function for *M*-estimators is then

$$\text{IF}(\mathbf{X},\mathbf{X}\hat{\beta}+\mathbf{e};T,F)=\left(\psi(\mathbf{e})\Big/\int\psi'(\mathbf{e})\,d\Phi\right)\cdot\left(\int\mathbf{X}'\mathbf{X}\,d\text{K}\right)^{-1}\mathbf{X}$$

where $\mathbf{e}=\mathbf{y}-\mathbf{X}\hat{\beta}$, Φ the normal distribution of the errors, and K is the distribution function of $\mathbf{x}_i$. There are two components of the IF, the influence of the residual (IR) and the influence of position (IP) such that

$$\text{IF}(\mathbf{X},\,\mathbf{X}\hat{\beta}\;+\mathbf{e};\,T,\,F)=\text{IR}(\mathbf{e};\,T,\,\Phi)\cdot\text{IP}(\mathbf{X};\,T,\,K)$$

For an estimator to have bounded influence, both the IR and IP must be bounded. For least squares, both the IR and IP are unbounded. The IR and IP for *M*-estimators are

$$\text{IR}=\left(\psi(\mathbf{e})\Big/\int\psi'(\mathbf{e})\,d\Phi\right)$$

$$\text{IP}=\left(\int\mathbf{X}'\mathbf{X}\,d\text{K}\right)^{-1}\mathbf{X}$$

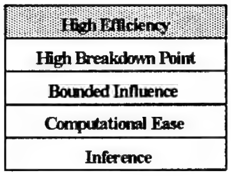
M-estimators have bounded IR if ψ is bounded, but unbounded IP. The class of estimators called GM-estimators can be configured such that both the IR and IP are bounded.

Computational ease: Considerations include the complexity and availability of the method used to calculate the estimates. This measure also considers the potential for convergence problems.

Inference: In order to test the adequacy of the estimation technique and choose the parameters which are significant in the model, hypothesis tests must be performed. These tests are more efficient if they are based on some, at least asymptotic, assumptions about the distribution of the estimator.

A graphic is displayed next to each robust technique discussed to quickly indicate the strengths and weaknesses of the method using these criteria. Criteria strengths are highlighted by shading.

2.2.1 L_1-norm or Least Absolute Values Estimation



High Efficiency
High Breakdown Point
Bounded Influence
Computational Ease
Inference

Many alternative estimators have been proposed for regression. One of these approaches came from Edgeworth (1887), improving a proposal of Boscovich (1757). He proposed the L_1-norm or least absolute values (LAV) regression estimator, which is determined by

$$\min \sum_{i=1}^{n} |e_i| \tag{2.7}$$

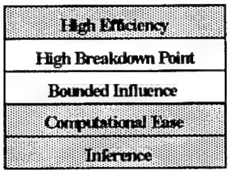
This approach attempts to minimize the sum of the absolute errors. The LAV estimator is commonly solved with linear programming methods. Unfortunately, the breakdown point of LAV regression is still no better than 0%. The LAV is robust to an outlier in the y-direction (unlike least squares). However, LAV regression does not protect against outlying x, where the effect of the leverage point is even stronger than on the least squares line. It turns out that when the leverage point is far enough away, the LAV line passes right through it. So a single erroneous point can totally offset the LAV estimator.

The L$_1$-norm and least squares (L$_2$-norm) are special cases of the L$_p$-norm regression problem. The objective in the general case is to

$$\min \sum_{i=1}^{n} \left| e_i \right|^p \tag{2.8}$$

where $1 \leq p \leq 2$. This approach has been considered by Gentlemen (1965), Forsythe (1972) and Sposito et al. (1977). Dodge (1984) suggested a regression estimator based on the convex combination of the L$_1$ and L$_2$ norms. All these proposals possess a zero breakdown point.

2.2.2 *M*-estimation



▨	High Efficiency
	High Breakdown Point
	Bounded Influence
▨	Computational Ease
▨	Inference

Huber (1973) introduced a class of estimators called "*M*-estimators". This method is the most popular of all robust estimators. The *M*-estimators are based on the idea of replacing the squared residuals by another function of the residuals $\rho(e)$, where ρ is a symmetric function with a unique minimum at zero.

$$\min_{\beta} \sum_{i=1}^{n} \rho(e_i) = \min_{\beta} \sum_{i=1}^{n} \rho(y_i - \mathbf{x}_i' \hat{\beta}) \tag{2.9}$$

M-estimators are maximum likelihood estimators where the function ρ is related to the likelihood function for an appropriate choice of the error distribution. Because the *M*-estimator is not scale invariant the minimization problem is modified by dividing the ρ function by a robust estimate of scale s, so the formula becomes

$$\min_{\beta} \sum_{i=1}^{n} \rho\left(\frac{e_i}{s}\right) = \min_{\beta} \sum_{i=1}^{n} \rho\left(\frac{y_i - \mathbf{x}_i' \hat{\beta}}{s}\right) \tag{2.10}$$

A popular choice for s is

$$s = \text{median } |e_i - \text{median}(e_i)| \ / \ 0.6745$$

The constant 0.6745 is used to make s an asymptotically unbiased estimator of σ and the sample actually arises from a normal distribution.

The least squares estimator is a special case of the $\rho()$ function where $\rho(u)=\frac{1}{2}u^2$. For a convex ρ, equivalence to (2.10) can be found by finding the first partial derivatives of (2.10) with respect to β and setting the result equal to **0**, as

$$\min_{\beta}\sum_{i=1}^{n}\psi\left(\frac{e_i}{s}\right)=\min_{\beta}\sum_{i=1}^{n}\psi\left(\frac{y_i-\mathbf{x}_i'\hat{\beta}}{s}\right)\mathbf{x}_i=\mathbf{0} \tag{2.11}$$

where $\psi(u)=\frac{\partial}{\partial u}\rho(u)$, resulting in the necessary condition normal equations. If $\psi(u)=u$, then (2.11) reduces to the normal equations yielding the least squares estimator. However, in the case of robust estimation, $\psi(u)$ is not linear so that (2.11) defines a nonlinear system of equations which requires an appropriate iterative technique.

The $\psi(u)$ function controls the weight given to each residual and is very important in determining the robust and efficiency properties of the estimator. Although a number of popular ψ-functions have been developed, they primarily belong to one of two categories: monotonic and redescending. The least squares ψ-function described earlier reveals its weakness in situations involving heavy-tailed distributions. The ψ-function is unbounded meaning large residuals receive heavy weights. The Huber function (Huber 1964), is an example of a monotone ψ-function defined as $\psi(u)=min(c_H,max(u,-c_H))$ which results in down-weighting the large residuals compared to least squares. Other ψ-functions redescend with increasing residual magnitude. The Tukey bisquare or biweight function (Beaton and Tukey 1974), is defined as $\psi(u)=u(1-(u/c_B)^2)^2$ for $|u|<c_B$ and 0 if $|u|>c_B$. The c terms in both equations refer to tuning constants chosen to achieve desired efficiencies. The values c_H=1.345 and c_B=4.685 for the

Huber and Tukey ψ-functions each achieve 95% efficiency compared to the least squares estimator in the model when the errors are actually normally distributed. For an excellent summary of different approaches to the ψ-functions, see Montgomery and Peck (1992).

Several methods have been proposed to solve the set of nonlinear equations (2.11) resulting from these popular robust influence functions. Let $w(u) = \psi(u)/u$ be defined as the weight function and let < > denote an n x n diagonal matrix. Assume we have a starting value β_o, then the three most discussed iteration schemes for solving (2.11) are:

$$\hat{\beta}_1 = \hat{\beta}_0 + s\left(\mathbf{X}' < \psi'\left(\frac{\mathbf{y}-\mathbf{X}\hat{\beta}_0}{s}\right) > \mathbf{X}\right)^{-1} \mathbf{X}'\psi\left(\frac{\mathbf{y}-\mathbf{X}\hat{\beta}_0}{s}\right) \tag{2.12}$$

$$\hat{\beta}_1 = \hat{\beta}_0 + s(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\psi\left(\frac{\mathbf{y}-\mathbf{X}\hat{\beta}_0}{s}\right) \tag{2.13}$$

$$\hat{\beta}_1 = \hat{\beta}_0 + \left(\mathbf{X}' < w\left(\frac{\mathbf{y}-\mathbf{X}\hat{\beta}_0}{s}\right) > \mathbf{X}\right)^{-1} \mathbf{X}' < w\left(\frac{\mathbf{y}-\mathbf{X}\hat{\beta}_0}{s}\right) > \left(\mathbf{y}-\mathbf{X}\hat{\beta}_0\right) \tag{2.14}$$

The first equation (2.12) is Newton's method, the second is called the H-algorithm which was proposed and discussed by Huber(1973) and Bickel (1975) and the third is iteratively reweighted least squares (IRLS) from Beaton and Tukey (1974).

The three procedures coincide in the normal distribution case. Newton's method has the strongest theoretical justification and provides the greatest adjustment on the first iteration. However, because it requires the derivative of the influence function it is difficult to implement in computer packages and is only defined when $\psi'(u) > 0$ for all u; that is when ψ' is strictly increasing.

The Huber method has desirable properties since the generalized inverse $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, need only be computed once and has the advantage of using only the unweighted design matrix. The drawbacks of this approach are that it is not as easy to use with existing ordinary least squares regression packages and it generally requires more iterations.

The IRLS method is flexible enough either to use an existing weighted least-squares algorithm, or if that is not available, to compute the square root of the weight function and use a standard least squares program for each step. This method generally converges noticeably faster than Huber's method, but more slowly than Newton's method. Additionally, the IRLS approach is more general as it can be applied to linear and non-linear models, whereas the Huber approach only applies to linear models.

Iteratively reweighted least squares (IRLS) is the most widely used nonlinear optimization technique in robust regression. A major reason for the widespread application of IRLS is that it can be used in an ordinary or weighted least-squares framework. This can be demonstrated by expressing (2.14) as

$$\mathbf{X'WX}\hat{\beta} = \mathbf{X'Wy} \tag{2.15}$$

where **W** is an n x n diagonal matrix of weights with diagonal elements w_1, w_2, ...,w_n given by

$$w_i = \frac{\psi\left[\left(y_i - \mathbf{x}_i'\hat{\beta}_0\right)/s\right]}{\left(y_i - \mathbf{x}_i'\hat{\beta}_0\right)/s} \tag{2.16}$$

The equation in (2.15) results in the usual weighted least squares normal equations. Thus, the one-step *M*-estimator can be found at convergence, where

$$\hat{\beta}_1 = (\mathbf{X'WX})^{-1}\mathbf{X'Wy} \tag{2.17}$$

At each iteration the weights are recomputed using the updated estimate of β. After the first iteration $\hat{\beta}_0$ is replaced by the updated estimate $\hat{\beta}_1$. Usually only a few iterations are required to achieve convergence.

One may be interested in the distributional properties of β . Huber (1981) showed that, under certain conditions, the asymptotic distribution of β is $N(\beta,\mathbf{V_A})$, where $\mathbf{V_A}$ is a function of σ^2, the ψ-function, and its derivative. Unfortunately, the finite sample distribution of β and its covariance matrix is not known. Holland and Welsch (1977) point out that one approach to robust inferential procedures based on β utilizes finite sample approximations to $\mathbf{V_A}$. They discuss several alternative finite sample estimates of the covariance matrix of β .

Concentrations of research have focused on the best technique for solving the system of equations. IRLS is the most popular approach, but subtleties in the approach are still unresolved. In each step of the iteration procedure, both the coefficients and the scale can be simultaneously reestimated. Convergence concerns arise when the scale estimate is reestimated. Some authors suggest iterating on scale (Rousseeuw and Leroy 1987; Street et al. 1988), while others suggest fixing the scale estimate (Hogg 1979; Green 1984). It is also very important to use a "good" starting value, one that is already sufficiently robust. Without this precaution one can easily end up in a local minimum that does not correspond at all to the expected robust solution. The calculation of bounded influence estimators presents similar problems.

M-estimators have taken the art of robust estimation to a higher, more applicable level. Vast amounts of research have been conducted constructing the ψ-functions so that the estimators are both robust **and** efficient. *M*-estimators are statistically more efficient than LAV regression, while at the same time they are robust with respect to outlying y. However, as will be discussed later in the section on bounded influence methods, *M*-estimators are not robust to **X**-space outliers. Also, their breakdown point is $1/n$ because of the effect of a single outlying observation.

2.2.3 *R*-estimation and *L*-estimation

High Efficiency
High Breakdown Point
Bounded Influence
Computational Ease
Inference

R-estimates are based on the ranks of the residuals. The idea of using these in multiple regression is attributed to Adichie (1967), Jaeckel (1972), and Jureckova (1971). The proposal of Jaeckel uses the rank R_i of the residual $e_i = y_i - \mathbf{x}_i\hat{\beta}$ in the objective function as

$$\min \sum_{i=1}^{n} a(R_i)e_i \tag{2.18}$$

where $a(R_i)$ is the scores function. Examples of scores functions are the Wilcoxon scores and median scores.

Several research efforts have focused on using a linear combination of order statistics to obtain a robust estimate called an *L*-estimator. The order statistics of a random sample of a continuous distribution are $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$, where $x_{(i)}$ is the i$^{\text{th}}$ order statistic. Bickel (1973) has proposed a class of one-step estimators for regression that depend on an initial estimate of β. Koenker and Bassett (1978) use analogs of sample quantiles for regression. The trimmed least squares of Ruppert and Carroll (1980) are also *L*-estimators.

The performance of *R*- and *L*- estimators have not been as good as the *M*-estimators for the regression problem (Heiler 1981). Montgomery and Peck point out that *L*-estimators do not always generalize clearly to multiple regression and both *R*- and *L*-estimates are more computationally difficult to obtain than *M*-estimates.

2.2.4 Least Median of Squares (LMS) Estimation

High Efficiency
High Breakdown Point
Bounded Influence
Computational Ease
Inference

Instead of using least sum of squares, which can also be interpreted as least squares on the *mean*, what about least squares on the *median*? This approach was first proposed by Hampel (1975, p. 380) and was later adopted and refined by Rousseeuw (1983, 1984). Rousseeuw proposed the least median of squares (LMS) estimator given by

$$\min_{\beta} med\ e_i^2 \tag{2.19}$$

This estimator can be robust with respect to outliers in both the x- and y-directions, but does not contain an influence function that is theoretically bounded in X-space. Its breakdown point can be as high as 50%, assuming the random subsample size is set appropriately. Unfortunately, the LMS estimator is not efficient relative to least squares when the errors are normal. Also, the computational effort involves evaluating all possible *p*-point subsets and using the estimate that produces the smallest median squared residual. This approach can result in the estimate being adversely effected by outliers. Because of its low efficiency, Rousseeuw and Leroy (1987) suggest using it for detecting outliers or as an initial stage estimator.

2.2.5 Least Trimmed Squares (LTS) Estimation

High Efficiency
▨ High Breakdown Point
Bounded Influence
Computational Ease
▨ Inference

The least trimmed squares (LTS) approach was developed also by Rousseeuw (1983, 1984) as a high efficiency alternative to LMS. The LTS estimator is given by

$$\min_{\beta} \sum_{i=1}^{h} (e^2)_{i:n} \tag{2.20}$$

where $(e^2)_{1:n} \le (e^2)_{2:n} \le \ldots \le (e^2)_{n:n}$ are the ordered squared residuals and h is the number of residuals included in the calculation. This approach is similar to least squares except the largest α squared residuals are not used (trimmed sum) in the summation, allowing the fit to avoid the outliers. This approach converges at a rate similar to the *M*-estimators. It is also equivariant and the breakdown point is 50% when *h*=*n*/2. According to Rousseeuw and Leroy (1987), the main disadvantage of LTS is the large number of operations required to sort the squared residuals in the objective function. Another challenge is deciding the best approach for determining the initial estimate.

2.2.6 *S*-estimators

▨	High Efficiency
▨	High Breakdown Point
	Bounded Influence
	Computational Ease
▨	Inference

This technique consists of a class of estimates based on the minimization of a robust *M*-estimate of the residual scale. They are defined by minimization of the dispersion of the residuals:

$$\min_{\beta} \ s\left(e_1(\beta),\cdots,e_n(\beta)\right) \tag{2.21}$$

with final scale estimate

$$\hat{\sigma} = s\left(e_1(\hat{\beta}),\cdots,e_n(\hat{\beta})\right) \tag{2.22}$$

The dispersion function $s\left(e_1(\beta),\cdots,e_n(\beta)\right)$ is found as the solution to

$$\frac{1}{n-p}\sum_{i=1}^{n}\rho\left(\frac{y_i-\mathbf{x}_i'\hat{\beta}}{s}\right)=K \tag{2.23}$$

The constant *K* may be defined as $E_{\Phi}[\rho]$, where Φ stands for the standard normal distribution. The function ρ must be such that the following conditions are met:

C1: ρ is symmetric and continuously differentiable, and $\rho(0) = 0$.

C2: There exists $c > 0$ such that ρ is strictly increasing on $[0, c]$ and constant on $[c, \infty]$.

The term *S*-estimator is used to describe this class of robust estimation because the scale statistic *s* is implicitly derived as an *M*-estimate of scale. These estimators have the characteristics of a high breakdown point (up to 50%), and equivariance (regression, scale and affine). *S*-estimates are also asymptotically normal with rate of convergence $n^{1/2}$.

The usual choice for the ρ function is

$$\rho_c(s) = \begin{cases} 3(s/c)^2 - 3(s/c)^4 + (s/c)^6 & \text{if } |s| \le c \\ 1 & \text{if } |s| > c \end{cases}$$

the derivative of which is the biweight ψ-function of Beaton and Tukey (1974) given by

$$\psi(s) = \begin{cases} s(1-s^2)^2 & \text{if} |s| \le 1 \\ 0 & \text{if} |s| > 1 \end{cases}$$

The breakdown point of *S*-estimators can be 50%, assuming a condition is satisfied relating the constant *K* with the ρ function such that

$$\text{C3: } \frac{K}{\rho(c)} = \frac{1}{2}$$

For example, for the case of K = E[ρ], the 50% breakdown is obtained with c = 1.548. The *S*-estimator breakdown point is also slightly affected by the sample size and number of model parameters. For the highest breakdown combination of *K* and ρ, the actual breakdown point is

$$bp_n^* = ([n/2] - p + 2)/n$$

The corresponding asymptotic (relative) efficiency for the Normal error model can be calculated for various breakdown combinations of *K* and *c*. Efficiency can be increased at the expense of decreases in breakdown point. Some situations may warrant this type of tradeoff and subsequent selection of appropriate values for *K* and *c*. Table 2.1 lists some alternatives to the high breakdown point constants. There is an obvious tradeoff of breakdown for efficiency in the case of *S*-estimators, which is the reason for the partial shading of the high breakdown and high efficiency blocks in the quick indication block at the start of this section. The tuning constants can be set for maximum breakdown or high efficiency. Some flexibility is available so that the estimator can be used either as a high breakdown initial estimate with higher efficiency than LMS or LTS, or as a moderate breakdown (25%), moderate efficiency (75.9%) estimator.

Table 2.1. Breakdown Point (*bp) and Asymptotic Efficiency (*ae*) of *S*-estimators Using Tukey's Biweight Function and Various Combinations of Constants *c* and *K***

*bp** (%)	*ae* (%)	*c*	*K*
50	28.7	1.548	0.1995
45	37.0	1.756	0.2312
40	46.2	1.988	0.2634
35	56.0	2.251	0.2957
30	66.1	2.560	0.3278
25	75.9	2.917	0.3593
20	84.7	3.420	0.3899
15	91.7	4.096	0.4194
10	96.6	5.182	0.4475

Source: Table 19 of Rousseeuw, P. J., and Leroy, A. M. (1987), *Robust Regression and Outlier Detection*, Wiley, N. Y.

S-estimators satisfy the same first-order necessary conditions as *M*-estimators. Considering the equations obtained by differentiating the ρ function objective equation, and by denoting ρ - K by χ, we look for $\left(\hat{\beta},\hat{\sigma}\right)$ which are the solution to the system of equations

$$\begin{cases} \frac{1}{n}\sum_{i=1}^{n}\psi(e_i(\hat{\beta})/\hat{\sigma})\mathbf{x}_i = \mathbf{0} \\ \frac{1}{n}\sum_{i=1}^{n}\chi(e_i(\hat{\beta})/\hat{\sigma}) = 0 \end{cases} \tag{2.24}$$

These normal equations are the same as those defining the *M*-estimator so it is hoped that the same approach could be made in finding their solution. Unfortunately, these equations cannot be used because there are multiple solutions (ψ is redescending). Iterations could result in a poor local solution especially in the case of high leverage points.

A resampling algorithm similar to the type necessary for LMS and LTS estimates is required to find *S*-estimates. The purpose of the algorithm is to provide a reasonable approximation to the solution to the minimization problem. The approximation usually shares the breakdown point and equivariance properties of the exact estimate. Refinements are often added to the resampling approximation in an effort to satisfy the necessary conditions.

2.2.7 Bounded Influence or Generalized *M*-estimators

▨ High Efficiency
High Breakdown Point
▨ Bounded Influence
▨ Computational Ease
▨ Inference

The *M*-estimator can successfully handle situations where the outliers in the response variable occur at points in the regressor space with low to moderate leverage. Outliers occurring outside the regressor space in either the response variable or independent variable direction at high leverage locations create problems not only for the least squares estimator, but for the *M*-estimator as well. In particular, *M*-estimators are vulnerable to points having a small residual with the corresponding leverage or influence on the regression equation being very large. These small residual, high leverage points could receive full weight under *M*-estimation.

The diagonal elements of the "hat matrix" $\mathbf{H{=}X(X'X)^{-1}X'}$, denoted h_{ii} are typically used as measures of leverage in regression. The h_{ii} is a standardized measure of the distance of a point x_i' to the centroid of the regressor space. The range of h_{ii} is $1/n \leq h_{ii} \leq 1$ and the average value of h_{ii} is *p/n*. Hoaglin and Welsch (1978) suggest that values of $h_{ii} > 2p/n$ can be considered high leverage points.

A robust technique that attempts to downweight the high influence points as well as large residual points is Generalized *M*-estimation (GM-estimation). The GM estimators are solutions to the normal equations formed by

$$\sum_{i=1}^{n} \pi_i \psi\left(\frac{y_i - \mathbf{x}_i'\hat{\beta}}{s\pi_i}\right)\mathbf{x}_i = \mathbf{0} \tag{2.25}$$

where, for appropriate values of π_i the GM-estimator can downweight outliers with high leverage points. The estimator described here was developed by Schweppe (see Hill 1977). The other main type of GM-estimator was proposed by Mallows (1975). The distinction between these two types of objective functions is that the Mallows estimator does not have the π-weights in the denominator of the ψ-function. Both types have the effect of downweighting leverage points, but the Schweppe weighting scheme tends to downweight only if the residuals are large. Krasker and Welsch (1982) describe a weakness in the Mallows estimator:

> Outlying points in the X space increase the efficiency of most estimation procedures. Any downweighting in X space that does not include some consideration for how the y values at the outlying observations fit the pattern set by the bulk of the data cannot be efficient.

They go on to say that the Schweppe estimator has the potential to overcome these efficiency problems.

IRLS can be used again to solve (2.25). At convergence, the GM-estimator can be written as

$$\hat{\beta}_{\text{GM}} = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{y} \tag{2.26}$$

where in this case the diagonal elements of $\boldsymbol{W}$ are the weights w_i defined as

$$w_i = \frac{\psi\left[\left(y_i - \mathbf{x}_i'\hat{\beta}_{\text{GM}}\right)/\pi_i s\right]}{\left(y_i - \mathbf{x}_i'\hat{\beta}_{\text{GM}}\right)/\pi_i s} \tag{2.27}$$

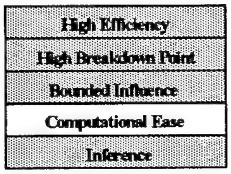
Several authors, including Krasker and Welsch (1982) suggest that the π_i take the form

$$\pi_i = \left[\left(1 - h_{ii}\right)/h_{ii}\right]^{1/2} \tag{2.28}$$

Several suggestions for the π-weights have been made that involve typical least squares outlier diagnostics, including DFFITS, which is used in (2.28) above. Other suggestions include studentized residuals, PRESS residuals or even Cook's D statistic. Each of these diagnostics measures leverage to some degree because each contains h_{ii} in their respective equation. Suggestions for the ψ-functions include various different *M*-estimate approaches such as Huber's *t* and Tukey's biweight. The research in this area is fairly new and some untried combinations of π-weights and ψ-functions could produce excellent estimators.

GM-estimators possess the same efficiency and asymptotic distributional properties as *M*-estimators. The breakdown point of the GM approach improves on the 1/*n* value of *M*-estimation, but is still not considered a high breakdown point estimator. The breakdown point is a function of the number of variables *p*, and is no greater than 1/*p*. This condition can lead to problems in models with many regressors. Also, both *M*-estimation and GM-estimation can be improved by starting with a good initial estimate. Advances in multi-stage GM-estimators that use high breakdown initial estimators, high breakdown π-weights, and modified convergence schemes result in final estimates with additional desirable properties. Some of these approaches are discussed in the following section.

2.2.8 Multi-stage GM-estimators



The discussion of robust estimators has clearly shown that no estimator has all of the desirable properties. Some of the methods have been proposed to obtain good initial estimators, while others reveal that they can be enhanced by a good initial estimate. The intent of multi-stage GM-estimators is to take advantage of these complementary needs. The purpose is to use different techniques in different stages so that the desirable properties of each technique can be combined. For example, if an LMS estimator can be effectively combined with a GM-estimator and the properties maintained, then an estimate could be developed that has high efficiency, high breakdown, bounds the influence, and has asymptotic distributional properties. Although this idea has been around for several years (Hampel et al. 1986; Rousseeuw and Leroy 1987; and Ronchetti 1987), only in the last few years have techniques actually been developed. Simpson et al. (1992) and Coakley and Hettmansperger (1993) both propose two-stage estimators that use high breakdown point initial estimates to generate good starting values and GM-estimation to increase the efficiency and bound the influence of the final estimate. Simpson et al. use an LMS initial stage and Mallows type GM objective function for the second stage estimator. Coakley and Hettmansperger propose an LTS initial estimate, followed by a Schweppe type bounded-influence estimator.

Both approaches use a one-step estimation method to solve the system of equations for the second stage estimate after finding the initial estimate $\hat{\boldsymbol{\beta}}_0$. They both use the Newton-Raphson method of (2.12). Simpson et al. state that one-step estimation inherits the breakdown point of the initial estimator and at the same time maintains the sample distribution of the *secondary* estimate. They say that IRLS inherits the asymptotic distribution of the *initial* estimate. More investigation is required here to determine the best approach to use in solving for the second stage estimate.

Coakley and Hettmansperger show that their estimator satisfies the goals of high breakdown, bounded-influence, and high efficiency. They also derive the asymptotic sampling distributions showing that the estimator is asymptotically normal, similar to the fully iterated GM-estimator.

This multi-stage approach to robust estimation clearly shows the most promise in terms of attaining desirable properties. Many different choices of estimators are available for each of the stages. The methodology discussed in Chapter 4 describes some of the possibilities.

2.2.9 MM-estimation

High Efficiency
High Breakdown Point
Bounded Influence
Computational Ease
Inference

Yohai (1987) introduced multi-stage estimators called MM-estimates, which combine high breakdown with high asymptotic efficiency. MM-estimates are computed using a three-stage procedure. The first step involves the computation of an initial estimate with high breakdown properties. Yohai suggests using the *S*-estimate for this initial estimator. The second stage is used to compute an *M*-estimate of the errors scale using the initial step *S*-estimate residuals. Lastly, in the third stage an *M*-estimate of the regression parameters based on an appropriate redescending ψ-function is computed.

Since Yohai's (1987) original proposal, refinements have been suggested by several authors including Ruppert (1992) and Yohai et al. (1991). The algorithm described below includes these refinements and the suggested implementation is provided by Marazzi (1993). This implementation includes the test for bias suggested by Yohai et al. which uses a student T test statistic to determine whether the bias in the final estimate may be unacceptably high and perhaps the initial estimate should be used for exploratory purposes.

The modified approach includes a three-step procedure beginning with an initial high breakdown *S*-estimate.

First define the *M*-estimate of residual scale to be the integral of the Tukey biweight ψ-function such that for any *c*, let

$$\chi_c(s) = \begin{cases} 3(s/c)^2 - 3(s/c)^4 + (s/c)^6 & \text{if } |s| \le c \\ 1 & \text{if } |s| > c \end{cases}$$

1. Compute the initial high breakdown coefficient and scale estimates. Calculate the refined resampling approximation $\hat{\beta}_0$ using subsamples of size p for the S-estimator $\hat{\beta}_*$ defined by

$$\hat{\beta}_* = \arg\min_\beta S(\beta)$$

where $S(\beta)$ is the solution to

$$\frac{1}{n-p} \sum_{i=1}^{n} \chi_{c_0}\left(\frac{y_i - \mathbf{x}_i' \hat{\beta}}{s} \right) = K \tag{2.29}$$

with c_0 = 1.548 and$K = \int \chi_{c_0}(s) d\Phi(s) = 0.5$. Let $\sigma_0 = S(\beta_0)$ be the associated estimate of σ and

$e_i = y_i - \mathbf{x}_i{}' \hat{\beta}_0$ for $i = 1, \ldots, n$.

2. Compute the final MM-estimate β_1 corresponding to the local minimum of .

$$Q_{MM}(\hat{\beta}) = \sum_{i=1}^{n} \chi_{c_1}\left(\frac{y_i - \mathbf{x}_i' \hat{\beta}}{\sigma_0} \right) \tag{2.30}$$

such that $Q_{MM}(\hat{\beta}_1) \le Q_{MM}(\hat{\beta}_0)$, where c_1 = 4.687 (see Ruppert 1992). Let $e_i = y_i - \mathbf{x}_i{}' \hat{\beta}_1$ for $i = 1, \ldots, n$.

3. Compute the final scale estimate.

Let $\sigma_1 = S(\beta_1)$ be the solution to

$$\frac{1}{n-p} \sum_{i=1}^{n} \chi_{c_0}\left(\frac{y_i - x_i' \hat{\beta}}{s} \right) = 0.5 \tag{2.31}$$

Yohai et al. (1991) suggest setting the number of subsamples (*NREP*) for the *S*-estimate in the first step to log(0.01) / log(1-2^{-p}) $\cong$ 4.6 * 2^p.

Unfortunately, although MM-estimators have a high breakdown and are asymptotically efficient, they do not necessarily have bounded influence, meaning that they may not perform especially well in the presence of high leverage points.

The high breakdown, high efficiency regression estimator described above with test for bias is a refinement to MM-estimation. The refinements are proposed by Yohai et al. (1991) and Ruppert (1992) and are implemented under the function supplied by ROBETH (Marazzi 1993).

2.3 Biased Estimation

The review of the literature in this section will not be nearly as broad as the robust topic review because a few of the techniques in biased-estimation are fairly well-known and proven to be quite successful. Some recent research describing slight modifications to the approaches will also be mentioned. The techniques associated with biased estimation that have been used by those modeling the combined influence-collinearity problem have mostly involved ridge or generalized ridge regression. Askin and Montgomery (1984) show that although some of the other techniques, including principal components regression and Stein shrinkage, can be effectively combined with robust estimation, they are consistently outperformed by ridge and generalized ridge. The most common approaches to ridge and generalized ridge regression will be briefly described. Most of this introductory information is contained in Montgomery and Peck, who present a thorough summary of biased estimation approaches.

2.3.1 Ridge Regression

Ridge regression is the most popular and commonly used method for dealing with multicollinearity. The objective is to reduce the size and variance of the least squares estimates by introducing a slight amount of bias. This approach was originally proposed by Hoerl and Kennard (1970a, b). The ridge estimator is determined by solving a modified version of the least squares normal equations. The ridge estimator, $\hat{\beta}_R$, is given by

$$\hat{\beta}_R = [\mathbf{X}'\mathbf{X} + k\mathbf{I}]^{-1}\mathbf{X}'\mathbf{y} \tag{2.32}$$

where $k \geq 0$ is called the biasing parameter and is selected by the analyst. The challenge in this approach is finding the appropriate selection of k. Many methods for choosing k have been proposed. The approach recommended by Hoerl and Kennard (1970a) is to choose k by inspection of the ridge trace. The objective is to select the smallest value of k in which the estimate of $\hat{\boldsymbol{\beta}}_R$ stabilizes. Other suggested approaches that are more analytical have been proposed by Hoerl and Kennard (1976), McDonald and Galarneau (1975), and Mallows (1973). If the analyst's primary purpose in developing a model is prediction, Montgomery and Friedman (1993) propose choosing k that minimizes PRESS$_R(k)$, which is the PRESS statistic calculated for the ridge estimator.

An import computational aspect of ridge regression is that the estimates may be found by using an ordinary least squares program and augmenting the standardized data. This approach gives

$$\mathbf{X}_A = \begin{bmatrix} \mathbf{X} \\ \sqrt{k}\mathbf{I}_p \end{bmatrix} \qquad \mathrm{y}_A = \begin{bmatrix} \mathbf{y} \\ \mathbf{0}_p \end{bmatrix} \tag{2.33}$$

where $\sqrt{k}\mathbf{I}_p$ is a p x p diagonal matrix with diagonal elements equal to $\sqrt{k}$. The associated ridge estimates are computed by

$$\hat{\beta}_R = [\mathbf{X}'_A\mathbf{X}_A + k\mathbf{I}]^{-1}\mathbf{X}'_A\mathbf{y}_A = [\mathbf{X}'\mathbf{X} + k\mathbf{I}_p]^{-1}\mathbf{X}'\mathbf{y} \tag{2.34}$$

This augmented matrix approach can be used effectively with iteratively reweighted least squares to form a combined biased-robust estimator.

2.3.2 Generalized Ridge Regression

Generalized ridge regression is an extension to ridge that was proposed by Hoerl and Kennard (1970a) that allows separate biasing parameters to be obtained for each regressor. Working with a model transformed to the space of orthogonal regressors simplifies the discussion. Assuming **T** is the orthogonal matrix of the eigenvectors of the **X** matrix, let **Z**=**XT** and $\boldsymbol{\alpha}=\mathbf{T}'\boldsymbol{\beta}$ so that $\boldsymbol{\alpha}$ become the transformed model coefficients. The generalized ridge coefficients become

$$\hat{\boldsymbol{\beta}}_{GR} = \mathbf{T}\hat{\boldsymbol{\alpha}}_{GR} \tag{2.35}$$

The mean square error is minimized by selecting $k_j = \sigma^2 / \alpha_j^2$. Several authors, including Hemmerle (1975) noted that choosing k_j in this fashion results in too much shrinkage. Montgomery and Peck suggest constraining the maximum increase in the residual sum of squares to between 1 and 20 percent. This approach was used by Askin and Montgomery (1984) in their analysis of augmented robust regression procedures.

2.4 Biased-Robust Estimation

As was mentioned previously, the study of estimation under the simultaneous problems of influence points and collinearity has not been researched nearly in as much depth as either of the single problems. In fact, it wasn't until 1973 that Holland introduced the first approach to estimating under the simultaneous conditions. Later, Askin and Montgomery (1980) introduced a family of estimators that combined robust *M*-estimation criteria with biased estimation constraints. Pfaffenberger and Dielman (1985) used a similar approach but replaced the *M*-estimate with LAV estimation. Lawrence and Marsh (1984), Askin and Montgomery (1984), and Pfaffenberger and Dielman (1990) compare alternative combinations of ridge regression and robust regression techniques. Askin and Montgomery, and Pfaffenberger and Dielman use designed experiments with Monte Carlo simulation, while Lawrence and Marsh use real data to predict fatalities in the US coal mining industry. Walker (1987) modified Askin and Montgomery's approach to allow bounded-influence estimators to be used instead of *M*-estimators, thus being able to better control the influence. Walker emphasizes the importance of applying these types of estimators in the combined problem by showing the potential effects of collinearity on robust estimators and also the effects of influence on biased estimators.

The approach suggested by Hogg (1979) and Askin and Montgomery (1980) was to apply some sort of robust estimation to a ridge regression model. The ridge estimator is first obtained by augmenting the least squares design and observation matrices with *p* additional rows. The robust ridge estimators are the solution to the problem

$$\min_{\beta} \sum_{i=1}^{n} \rho(y_i - \mathbf{x}_i^{'} \hat{\beta}) \tag{2.36}$$

$$\text{subject to } \hat{\beta}\hat{\beta} \le d^2$$

where the objective function is the classic *M*-estimator described previously. The solution to this problem can be obtained by IRLS where the weights on augmented observations are fixed at 1.0. The resulting estimator becomes

$$\hat{\beta} = [\mathbf{X}'\mathbf{W}^{\mathrm{k}}\mathbf{X} + k\mathbf{I}]^{-1}\mathbf{X}'\mathbf{W}^{\mathrm{k}}\mathbf{y} \tag{2.37}$$

where $\mathbf{W}_k = diag(w_1^k, w_2^k, \ldots, w_i^k)$ and $w_i^k = \psi(e_i^k / s) / (e_i^k / s)$. The weighting matrix is now a function of the shrinkage parameter *k*. Sensitivity to influential observations is a problem because *M*-estimators are used.

A natural extension of augmented robust estimators are augmented bounded-influence estimators (Walker 1987). The estimator in this case is the solution to

$$\min_{\beta} \sum_{i=1}^{n} \rho(\frac{y_i - \mathbf{x}_i'\hat{\beta}}{\pi_i s})\pi_i \tag{2.38}$$

subject to $\hat{\beta}\hat{\beta} \le d^2$

where the objective function is now the bounded-influence estimation approach. The estimator is found using (2.37) above, but in this case the weights are found by applying the bounded-influence approach where $w_i^k = \psi(e_i^k / \pi_i s) / (e_i^k / \pi_i s)$. The weights in this case are not fixed, but are functions of the shrinkage parameter k. Walker suggested using the DFFITS measure for the π-weights and he tested two variations of the ψ-function. A monotonic function (Huber's t) and a redescending function (Tukey's biweight) were compared.

Testing other variations of the Walker approach would be worthy of additional research. Emphasis could be placed on enhancing the robust estimation portion by focusing on the initial estimate and robust estimates leverage for the π-weights. Various combinations of ψ-functions and tuning constants could also be tested and analyzed for possible performance improvements over current approaches. Some alternative biased-robust estimation sequences could be investigated. The modified sequences may involve first using a robust estimator as the initial estimates for the biasing parameter. Some of these suggestions are implemented in the following chapters in an effort to improve estimation accuracy over current methods across a variety of scenarios.

Chapter 3

Outlier Characterization and Existing Robust Technique Comparison

3.1 Introduction

The situation often arises in the practice of fitting regression models to data that one or more of the observations is an aberration; that is, it does not follow the pattern of the majority of the data. This observation, called an outlier, may be obvious if we are working in small dimension problems that accommodate data plots. Even if the outlier is located, it may also be a valid observation worthy of keeping in the model. The traditional approach to regression, least squares, works well under most situations, but not if outliers are present. Least squares is influenced too heavily by outliers, to the point that the resulting estimation may be of no meaning for the bulk of the data.

Fortunately, a number of regression methods, called *robust* regression techniques, have been developed that are less sensitive to outliers than least squares. They offer a compromise between using least squares and omitting a potentially valid observation. There are a large number of robust regression alternatives and no technique is clearly superior to all the others. One reason for the lack of a superior technique is that there are so many types of outlier scenarios, each requiring a different set of robust fitting skills. A technique that fits well regardless of the outlier or nonoutlier condition would be an attractive commodity, especially for those analysts who use regression routinely and deal with more than two regressor variables.

In order to award a robust technique the title of "most robust", an understanding of the different types of outlier problems that identify vulnerabilities in various robust methods must first be researched and understood. Some of the outlier factors may be the percentage of data which are outliers, the location of the outlier in the regressor space, the location in the response direction, and whether the outliers are grouped or clustered. The purpose of this chapter is to perform a series of screening experiments to determine which realistic outlier scenarios cause differences in robust estimation behavior. A secondary objective of performing these screening experiments is to trim the list of candidate robust techniques for the "most robust" award by evaluating the performance of several methods.

The study of outlier characteristics will consist of two experiments designed in a sequential fashion so that the results of the first experiment can be used in deciding the best design for the second experiment. A candidate set of robust methods representative of the various types of estimators is used in the first experiment. The techniques will then be revised if necessary before the second experiment. The purpose of the first experiment, referred to as Pilot 1, is to study the performance of robust methods on datasets using various outlier locations and magnitudes. Outliers can be located in either interior **X**-space points or exterior **X**-space points. Exterior points are located far from the rest of the points in the regressor space and are also called high leverage points because they have a tendency to pull a least squares estimate in their direction. Outlier magnitude refers to the distance a point lies from a regression line fit to the nonoutliers. These two outlier factors are the focus of Pilot 1. Pilot 2 will consider varying the degree of leverage in the exterior points and consider datasets with variable magnitude outliers.

3.2 Desirable Properties and Robust Methods

Robust techniques are evaluated and compared with each other in terms of their ability to properly fit clean data as well as various outlier configurations. The properties of efficiency, breakdown and bounded influence are used to define technique capability in a theoretical sense. Efficiency refers to the expected performance of a robust technique relative to the performance of least squares, on clean data. High efficiency techniques are desired. Breakdown is the percentage of outliers present in the data before the technique's parameter estimates are meaningless. For instance, least squares has a breakdown of $1/n$, indicating that only one outlier can render the estimates useless. High breakdown is preferred and some robust techniques have the maximum possible breakdown point of $n/2$ or 50%. Fifty percent breakdown means that up to half of the observations can be discrepant and the estimator will still provide useful information. The third property, bounded influence, is based on the characteristic of least squares that allows points further out in the regressor space to exhibit greater influence on the parameter estimates. The purpose of bounding the influence is to reduce this exterior point influence, which can be especially important if these points are outliers.

The robust methods examined in Pilot 1 are techniques representing different classes of estimators based on these desirable properties. Each of the methods used are among the most respected in their category. Two high efficiency methods are used, *M*-estimation, and Least Absolute Value (LAV) estimation. Two high breakdown methods are evaluated, Least Median of Squares (LMS) and Least Trimmed Sums of Squares (LTS). Both methods have a 50% breakdown, but low efficiency relative to least squares. The other two robust methods are multiple property estimators, MM-estimation and Generalized *M*-estimation (GM-estimation). While the first four methods are single stage techniques, MM and GM-estimation are two-stage robust methods consisting of an initial robust estimate used to obtain a good starting point followed by a final robust estimate. The goal of multiple stage estimators is to obtain a final estimate that has two or more desirable properties. MM-estimation is a multi-stage technique that combines an initial *S*-estimate with a final *M*-estimate, resulting in an estimator that is both high efficiency and high breakdown. GM techniques are also multi-stage. The initial estimate is sometimes a high breakdown method, followed by a bounded influence final estimate. The technique tested in this experiment uses a most B-robust initial estimate which is actually an LAV estimator weighted by measures of leverage, and has a breakdown of $1/p$. This GM technique then has high efficiency and bounded influence. The measures of leverage used in the final estimate are *M*-estimates of covariance which also have a breakdown of $1/p$. Each of the methods used in Pilot 1 are listed in Table 3.1. along with comments on their desirable properties.

Table 3.1. Pilot 1 Experiment Regression Techniques and Associated Properties

Technique	Efficiency	Bounded Influence?	Breakdown
Least Squares (LS)	-	No	$1/n$
M-estimation (M)	High	No	$1/n$
Least Absolute Value (LAV)	High	No	$1/n$
Least Median of Squares (LMS)	Low	Yes	$n/2$
Least Trimmed Squares (LTS)	Medium	Yes	$n/2$
GM-estimation (GM)	High	Yes	$1/p$
MM-estimation (MM)	High	No	$n/2$

3.3 Pilot 1

The goal of Pilot 1 is to identify the appropriate arrangement of outliers in datasets so that robust techniques can be adequately and fairly tested for the desirable aspects of efficiency, breakdown and bounded influence. Outlier location and magnitude are the primary determinants of bounded influence and efficiency. A number of aspects of the dataset will be fixed so that the elements of outlier location and magnitude can be assessed. The fixed factors are listed in Table 3.3. The dataset will consist of a basic orthogonal fractional factorial design containing six regressor variables. The fractional factorial component will be a 2^{6-1} design consisting of ±1's representing 32 observations. The basic design is then augmented with eight axial points bringing the total sample size to 40. The axial points maintain the orthogonality of the columns and provide the high leverage point capability. Moving these points out on the axes allows the designer to control the amount of leverage. For both pilot experiments, the axial point distances are all equal and set to a value of 9.5. This distance was determined based on evaluations of several leverage measures. The leverage tests were designed to identify axial distances sufficiently far from center such that robust measures of leverage classified the axial points as outlying or exterior points in **X**-space. The techniques used to measure leverage were the hat diagonals and the robust distances using the minimum volume ellipsoid (MVE). Even though the hat diagonals (h_{ii}) are not considered robust, the leverage points in this dataset are symmetrically positioned the same distance from the **X**-space centroid so that the hat matrix estimate of location is not influenced by the high leverage points. Cutoff values have been proposed for these two diagnostics. Belsley, Kuh, and Welsch (1980) suggest h_{ii} values higher than $2p/n$ as indicators of high leverage. Rousseeuw and Leroy (1987) propose comparing the squared robust distances to a χ^2 statistic with p-1 degrees of freedom and an $\alpha = 0.025$ level of significance. Table 3.2. shows the measure of leverage results for the six-variable dataset with axial points located 9.5 units from the center. The Windows version of S-PLUS is used in conjunction with the software library ROBETH (Marazzi 1993) to compute these measures and to perform all the technique development and evaluation.

Table 3.2. Measures of Leverage for the High Leverage Points in Pilot 1

Observation	Robust Distance Squared	Hat Diagonals
Cutoff	*14.4*	*0.35*
33	36.7	0.45
34	31.9	0.45
35	31.9	0.74
36	36.7	0.74
37	36.7	0.74
38	64.2	0.74
39	31.9	0.45
40	36.7	0.45

The design responses **y** are determined using the form of the linear model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{3.1}$$

where **X** is the fractional factorial design matrix with specified axial points, $\boldsymbol{\beta}$ is the vector of known coefficients and $\boldsymbol{\varepsilon}$ is the vector of random errors which for these studies are random variates drawn from the standard normal distribution. The values of the true model coefficients $\boldsymbol{\beta}$ is set equal to 4.0 for each of the six regressors and to 0.0 for the intercept. The resulting signal-to-noise ratio of the model is then

$$SN = \Sigma \beta_i^{\,2} / \sigma_e^2 = 96 : 1$$

The outliers in each replicate run are formed by taking the sign from the N(0,1) random variate draw and changing the magnitude to the value specified by the design. The total number of outliers in each configuration is six, meaning that 15% of the sample points are outliers.

Table 3.3. Fixed Factors for Pilot 1 Experiment Design

Number of variables	6
Sample Size	Design $2^{6-1} = 32 + 8$ axial = 40
Outlier Percentage and Number	15%, 6
Number of Leverage Points	8
Coefficient Magnitude	4
Signal-to-Noise Ratio	96:1

As the purpose of this study states, the goal is to determine the impact of outlier magnitude and location on robust regression performance. The outlier magnitude and location factors are varied in the experimental design by using eight levels of outlier magnitude and three levels of outlier location (Table 3.4). For a particular experimental run, all of the outliers will have the same magnitude depending on the level. For instance, for the 8σ, interior and exterior **X**-space configuration, three outliers are generated in the interior **X**-space (from the 32 cube points) and three outliers are likewise created in the exterior **X**-space by assigning their $e_i = \text{sign}(N(0,1)) * 8.0$.

Table 3.4. Variable Factors for Pilot 1 Experiment Design

Factor	Levels
Outlier Magnitude	4σ, 6σ, 8σ, 10σ, 12σ, 14σ, 16σ, and 18σ
Outlier Location	Interior **X**-space, Interior and Exterior **X**-space, Exterior **X**-space

The eight levels of outlier magnitude and three levels of outlier location result in 24 separate treatment combinations that are available. All treatments combinations are performed and studied to gain the most insight possible regarding these two factors. The performance measures for each technique and each treatment combination is the mean square error of estimation (MSEE), which evaluates the deviation between a regression technique estimate and the true or known model coefficients.

$$\text{MSEE} = (\hat{\beta}_R - \beta)'(\hat{\beta}_R - \beta)$$

where $\hat{\beta}_R$ is a vector of regression technique parameter estimates and β is the vector of true model coefficients.

To increase the confidence level in the estimate of the MSEE, each treatment combination is replicated a number of times using different error random variate vectors. The statistic compiled from the replicates is the *average* MSEE or AMSEE. The AMSEE is computed by comparing the technique's regression parameter estimates to the true parameter values using

$$\text{AMSEE} = \text{mean}\left[\left(\hat{\beta}_R - \beta\right)'\left(\hat{\beta}_R - \beta\right)\right] \tag{3.2}$$

The number of replicates used in this study is 30, which results in a small MSEE sample standard deviation. In addition to the AMSEE, which evaluates a technique relative to the true model, one may also be interested in the relative improvement a robust technique provides over least squares estimation. A measure that provides a direct comparison to least squares is called the average mean square inefficiency ratio (AMSIR).

$$\text{AMSIR} = \frac{\text{mean}\left[(\hat{\beta}_R - \beta)'(\hat{\beta}_R - \beta)\right]}{\text{mean}\left[(\hat{\beta}_{LS} - \beta)'(\hat{\beta}_{LS} - \beta)\right]} \tag{3.3}$$

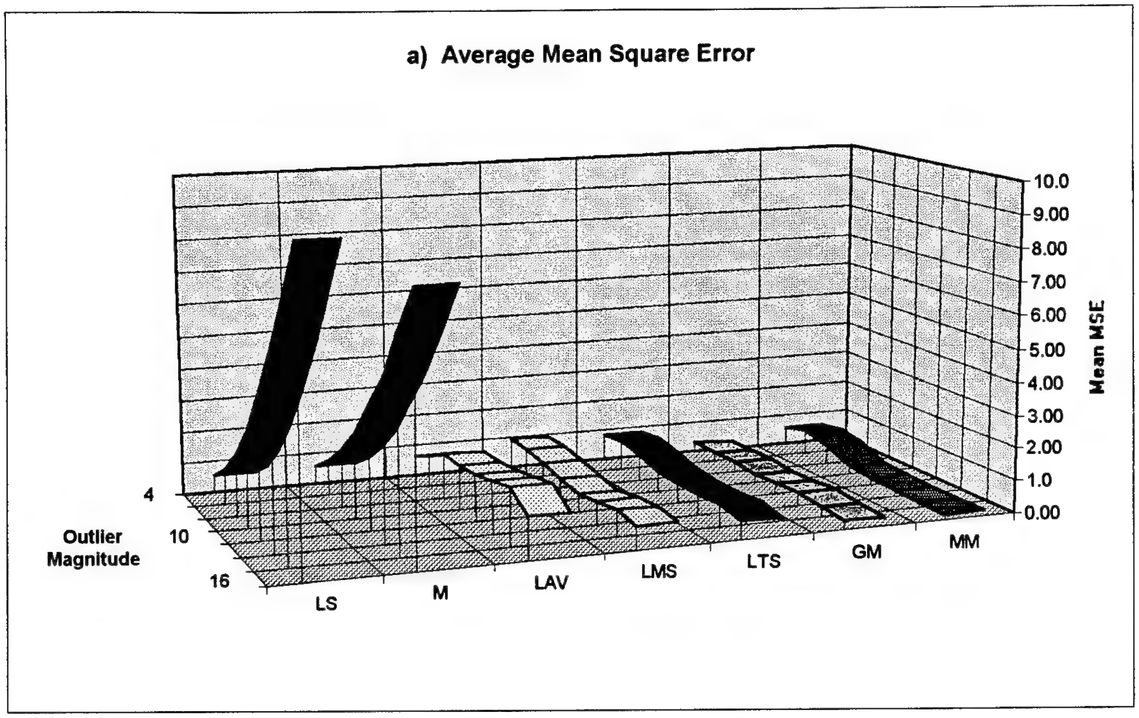
This measure consists of the ratio of the robust technique AMSEE to the least squares AMSEE. AMSIR values less than one indicate an improvement over least squares. Lower AMSIR values are desired.

3.3.1 Pilot 1 Individual Technique Performance

The results of the experiment will be discussed in terms of each technique's performance. Table 3.5 lists the AMSEE values for each technique in each experimental run. Recall that each AMSEE value averages performance over 30 model estimations. The strengths and weaknesses of each technique is listed along with comments on any unusual observations. Some general comments on the experiment are provided in the final section. The performance of the techniques in AMSEE and AMSIR relative to each other is displayed in Figures 3.1, 3.2, and 3.3.

Table 3.5. Pilot 1 Results in Average Mean Square Error of Estimation (AMSEE)

Interior **X**-space Outliers							
Outlier Magnitude	LS	M	LAV	LMS	LTS	GM	MM
4	0.17	0.16	0.15	0.88	0.61	0.28	0.20
6	0.64	0.15	0.23	0.51	0.55	0.43	0.20
8	0.60	0.13	0.20	0.53	0.66	0.34	0.13
10	0.99	0.09	0.13	0.38	0.47	0.26	0.09
12	1.42	0.09	0.17	0.46	0.35	0.28	0.09
14	2.40	0.09	0.15	0.46	0.50	0.36	0.11
16	2.60	0.08	0.15	0.48	0.46	0.29	0.09
18	3.35	0.10	0.18	0.40	0.41	0.33	0.09
Interior and Exterior **X**-space Outliers							
Outlier Magnitude	LS	M	LAV	LMS	LTS	GM	MM
4	0.27	0.32	0.31	0.81	0.76	0.28	0.38
6	0.61	0.44	0.40	0.84	0.70	0.30	0.39
8	1.05	0.53	0.48	0.74	0.68	0.30	0.27
10	1.64	0.39	0.52	0.49	0.50	0.27	0.19
12	2.54	0.36	0.51	0.56	0.48	0.30	0.14
14	3.17	0.28	0.40	0.57	0.45	0.27	0.13
16	4.46	0.17	0.58	0.60	0.60	0.30	0.13
18	5.98	0.13	0.54	0.66	0.57	0.32	0.12
Exterior **X**-space Outliers							
Outlier Magnitude	LS	M	LAV	LMS	LTS	GM	MM
4	0.53	0.56	0.67	0.94	0.85	0.30	0.57
6	1.09	1.08	0.98	0.92	0.97	0.30	0.76
8	1.92	1.83	1.36	0.85	0.85	0.33	0.61
10	2.98	2.67	1.39	0.68	0.62	0.28	0.24
12	4.35	3.83	1.46	0.55	0.57	0.30	0.19
14	5.99	4.99	1.62	0.71	0.61	0.37	0.19
16	7.64	6.47	1.81	0.83	0.70	0.34	0.17
18	9.43	8.01	1.30	0.73	0.55	0.28	0.16



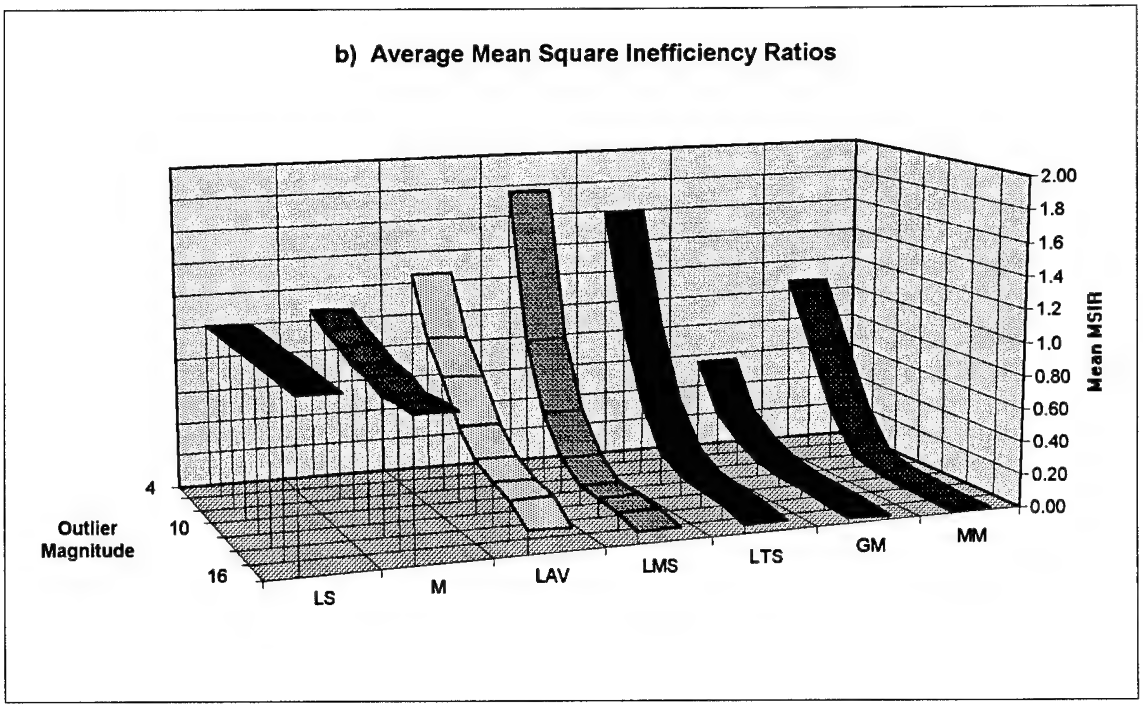
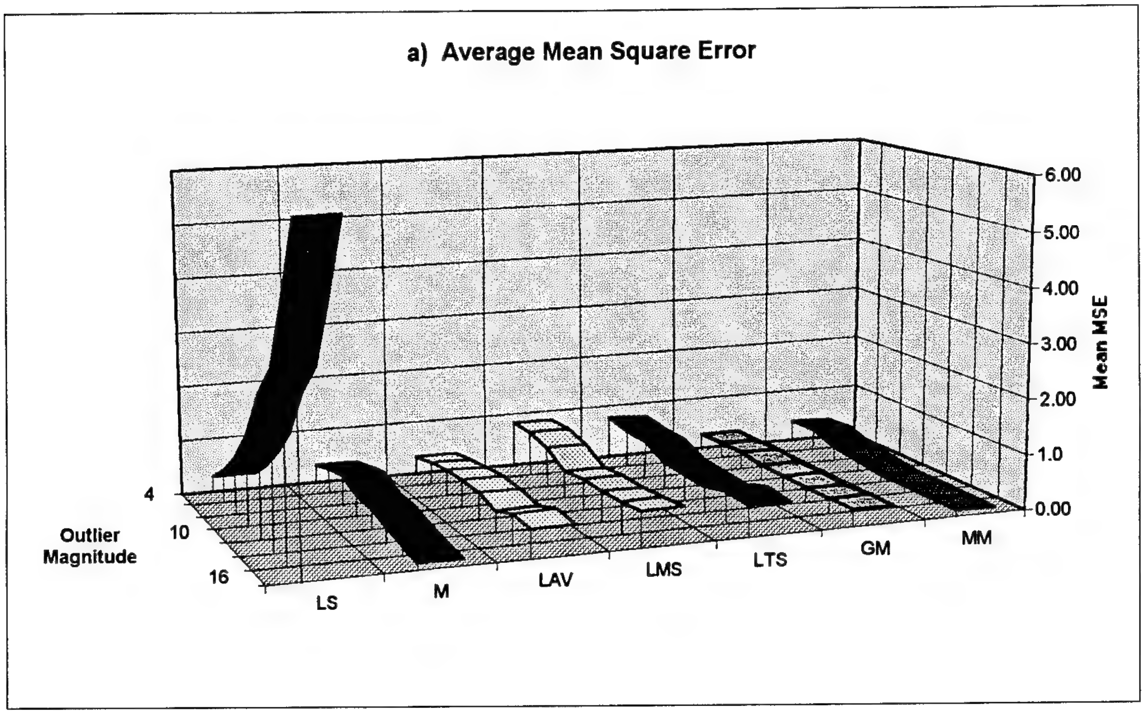


Figure 3.1. Pilot 1 Regression Technique Performance for Interior X-space Outliers in a) AMSEE and b) AMSIR



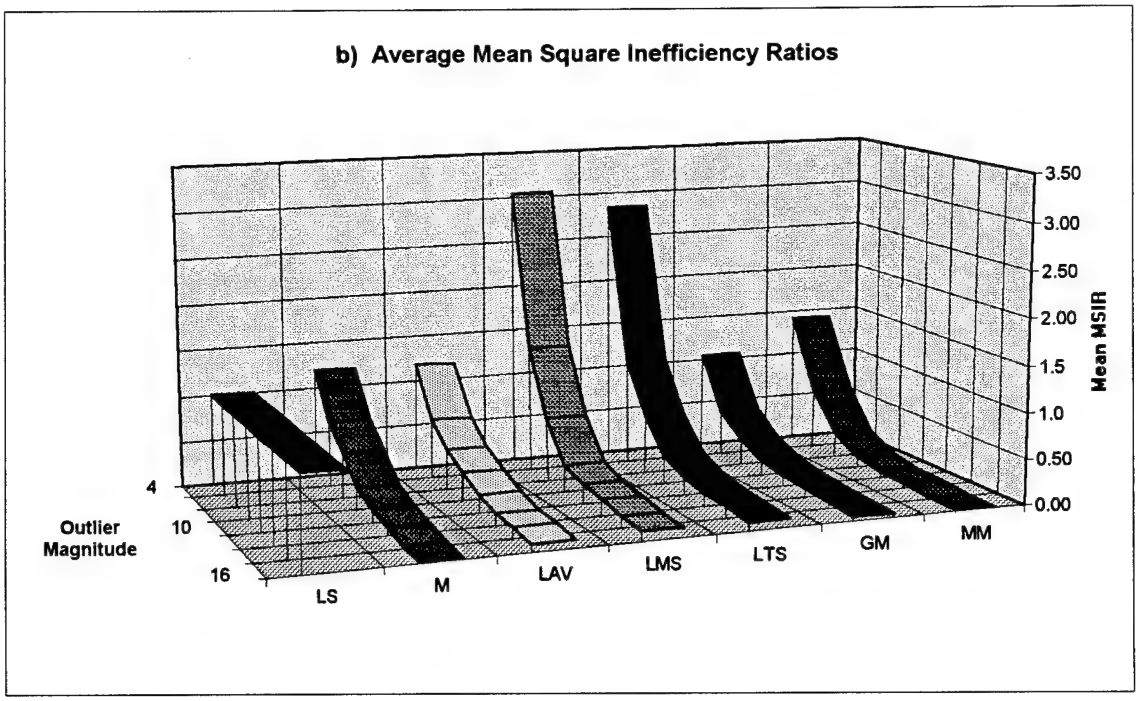
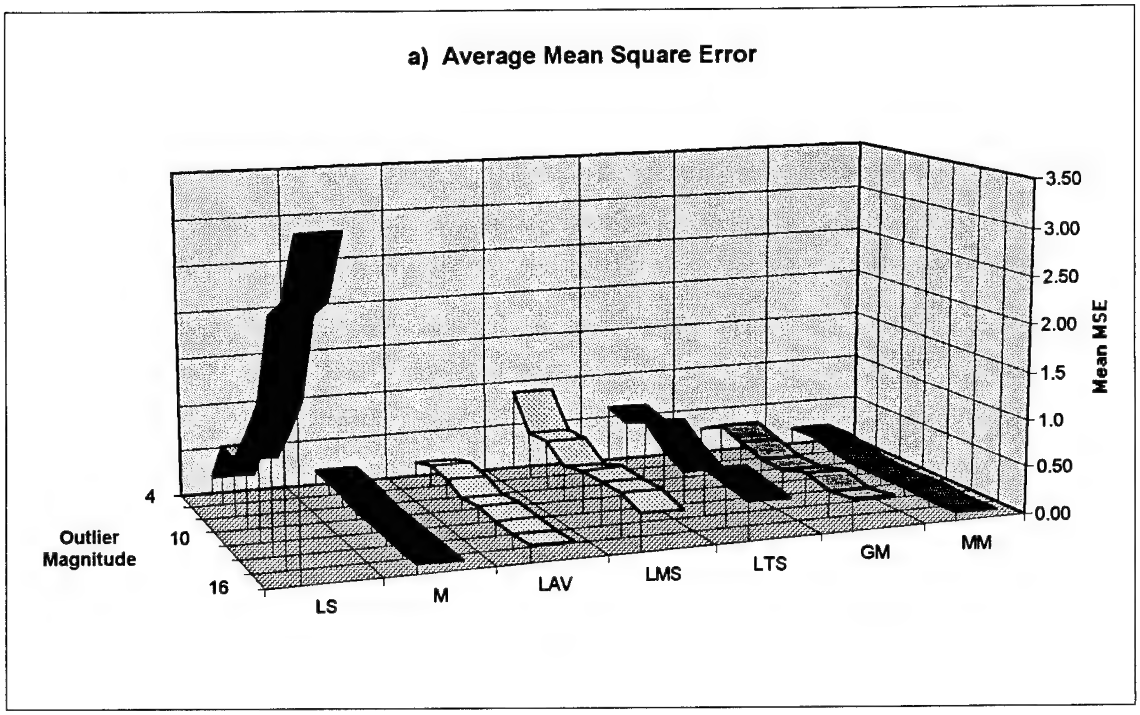


Figure 3.2. Pilot 1 Regression Technique Performance for Interior and Exterior X-space Outliers in a) AMSEE and b) AMSIR



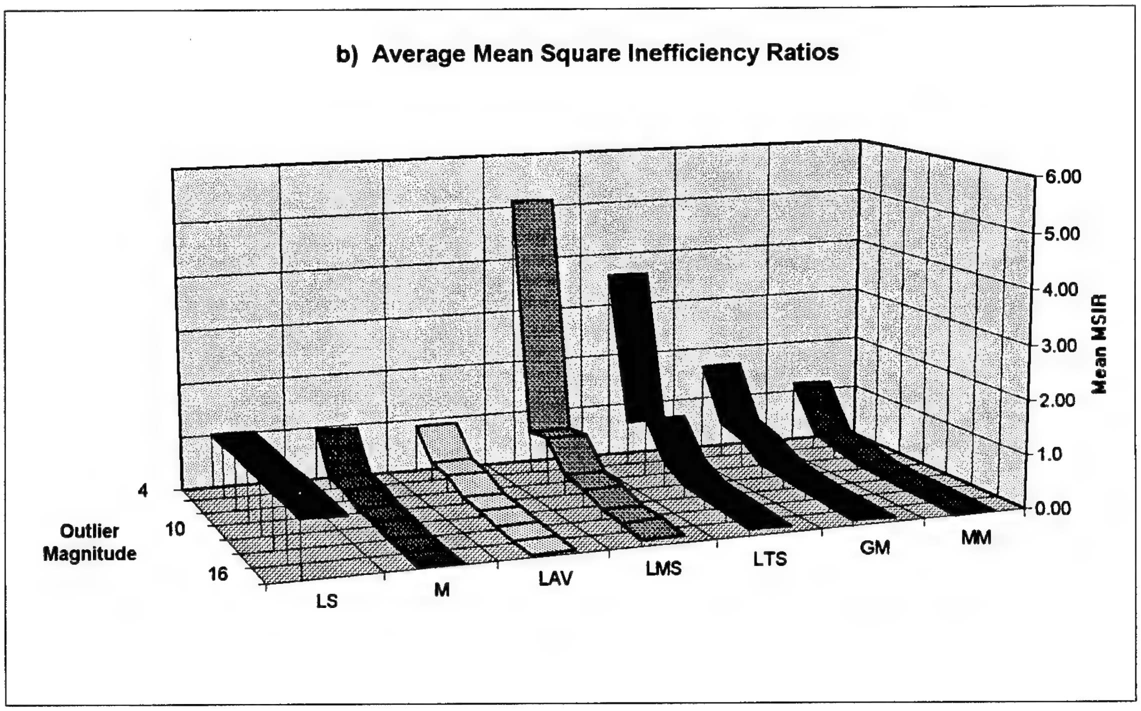


Figure 3.3. Pilot 1 Regression Technique Performance for Exterior X-space Outliers in a) AMSEE and b) AMSIR

Least Squares (LS)

- Regardless of location, LS performs worse (higher AMSEE) with increasing outlier magnitude
- Regardless of magnitude, LS performs worse as the percentage of exterior **X**-space outliers increases
- LS models have significantly higher MSEE values when the majority of the outliers have the same sign

M-*estimation (M)*

- Performs well (1st or 2nd best) in interior **X**-space outlier conditions; AMSEE decreases with increasing outlier magnitude
- Performance decreases relative to the other techniques as exterior **X**-space outliers are introduced; performs above average for exterior and interior **X**-space outliers and performs terribly for strictly exterior **X**-space outliers

Least Absolute Value (LAV)

- Performs above average (mostly 3rd) in interior **X**-space condition; AMSEE values do not depend on outlier magnitude
- Performs average in exterior and interior **X**-space condition; slight AMSEE increase with outlier magnitude
- Performs 2nd from last versus exterior **X**-space; AMSEE holds fairly steady though versus outlier magnitude

Least Median of Squares (LMS)

- In general, LMS shows low efficiency for each outlier location configuration (the 4σ outlier magnitude case)
- Has overall moderate performance versus other robust techniques
- Performs fairly well relative to the other techniques in the exterior **X**-space condition (some 3rd place finishes)
- Regardless of location, displays slight AMSEE decreases with increasing outlier magnitude

Least Trimmed Sum of Squares (LTS)

- Has slightly better efficiency (4σ scenario) than LMS, but still poor relative to the other robust methods
- Performs fairly well relative to the other techniques in the exterior **X**-space condition
- Regardless of location, LTS displays slight AMSEE decreases with increasing outlier magnitude

GM-estimation (GM)

- Most consistent performer, regardless of outlier magnitude; 3D surface plot (Figure 3.6a) of AMSEE versus outlier magnitude and outlier location is low and flat, meaning no weak areas
- Relative to other techniques, GM-estimation performs average in interior **X**-space, well in interior and exterior **X**-space, and well in exterior **X**-space

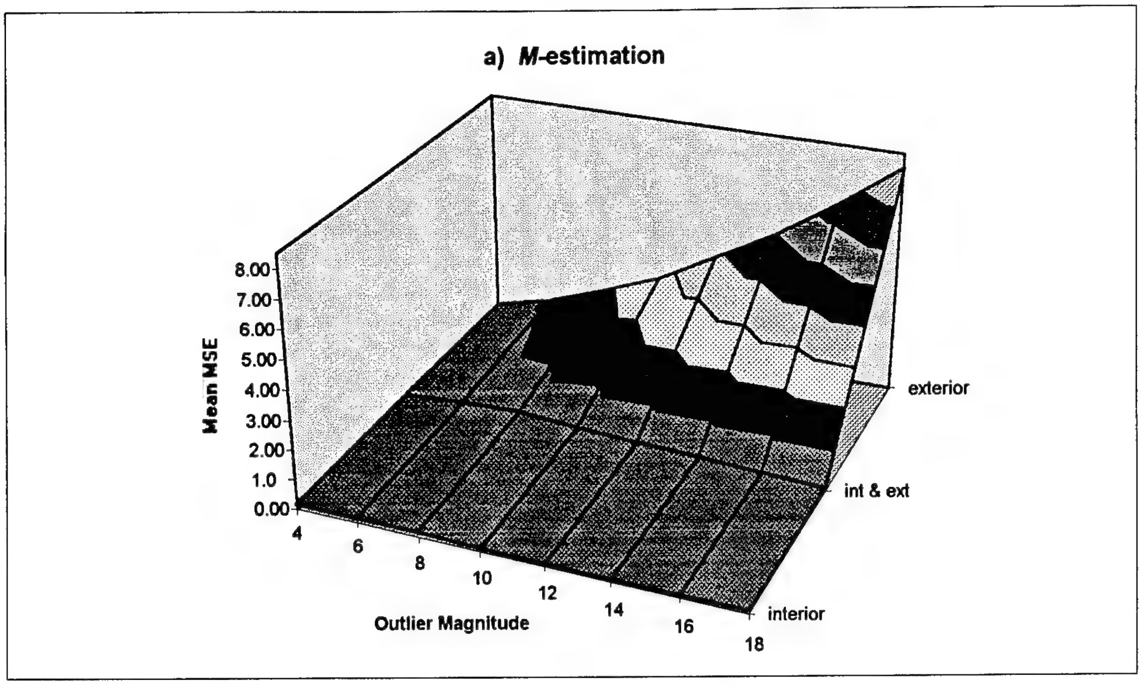
MM-estimation (MM)

- Best performer overall (1st or 2nd in all factor levels, except the 4σ case)
- Although MM-estimation does not theoretically bound influence, it performs very well in high leverage outlier situations
- Only area with slightly higher AMSEE was low magnitude (4σ, 6σ, and 8σ) **X**-space outliers

3.3.2 Pilot 1 Overall Performance

The overall performance of various methods or classes of methods is discussed. The performance of these techniques is displayed in the outlier magnitude and outlier location dimension in Figures 3.4, 3.5, and 3.6.

- High efficiency estimators perform well in the 4σ case, which is an adequate test for efficiency
- High breakdown estimators are not sufficiently tested for high breakdown (only 15% outliers in this study)
- GM-estimators perform well in situations involving high leverage outliers
- All robust techniques perform well relative to LS (low AMSIR); the single exception was *M*-estimation under the exterior **X**-space condition
- AMSEE values tend to decrease with increasing outlier magnitude, perhaps because outliers receive increasingly less weight to the point where they are dropped from the model



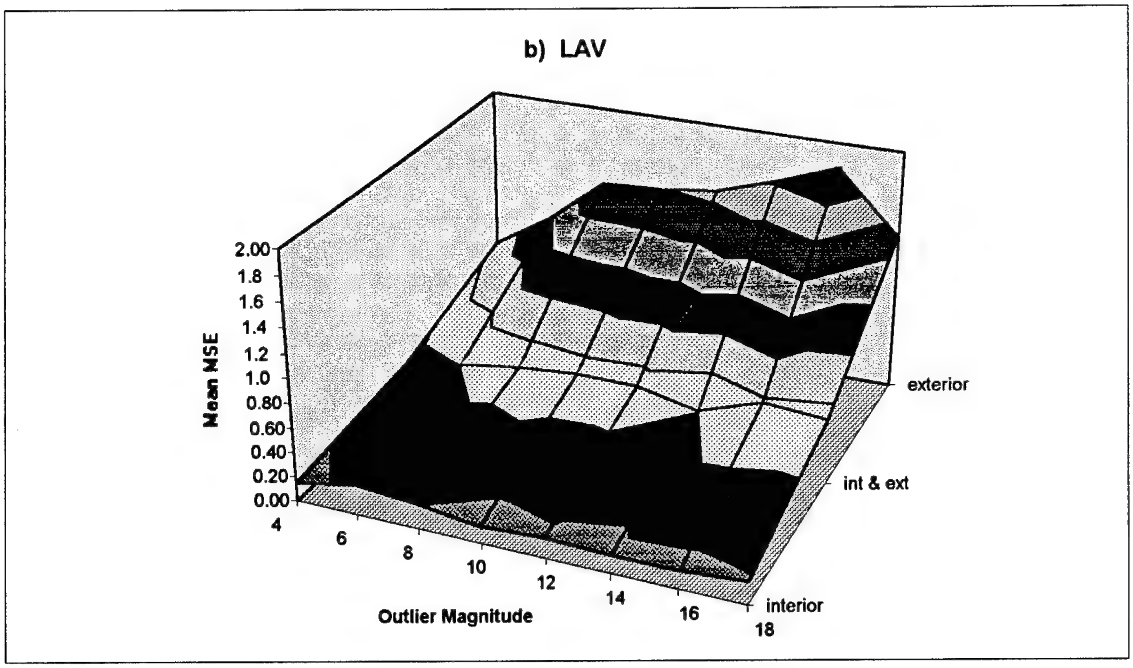
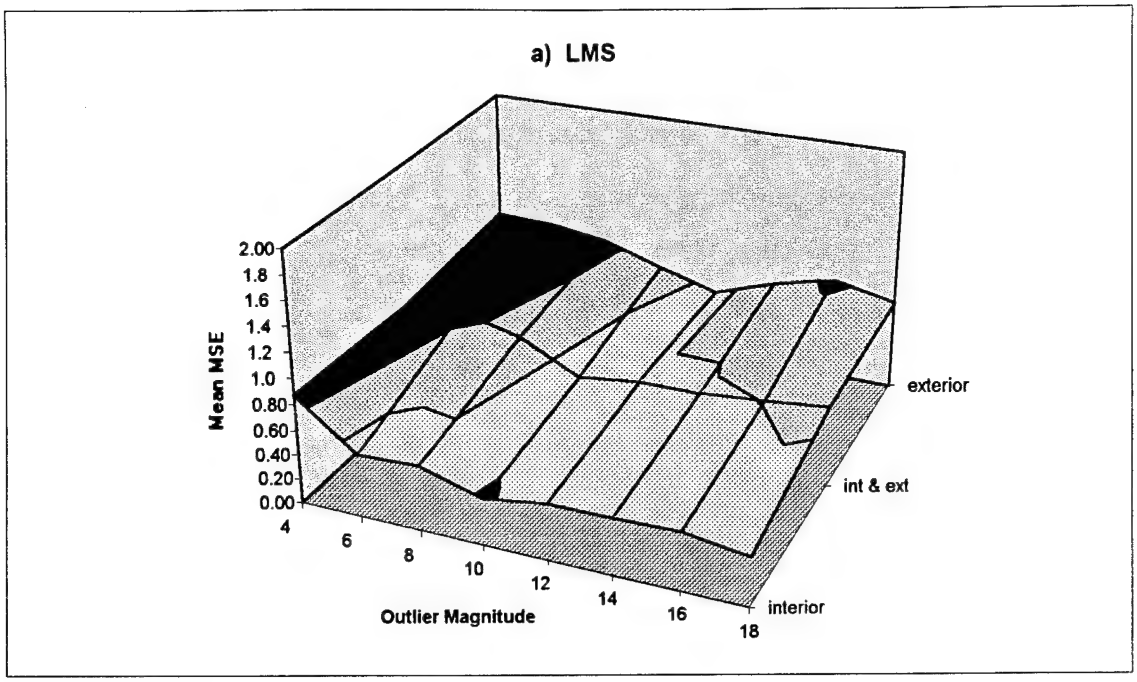


Figure 3.4. Pilot 1 Robust Regression Technique Overall Performance a) *M*-estimation and b) LAV



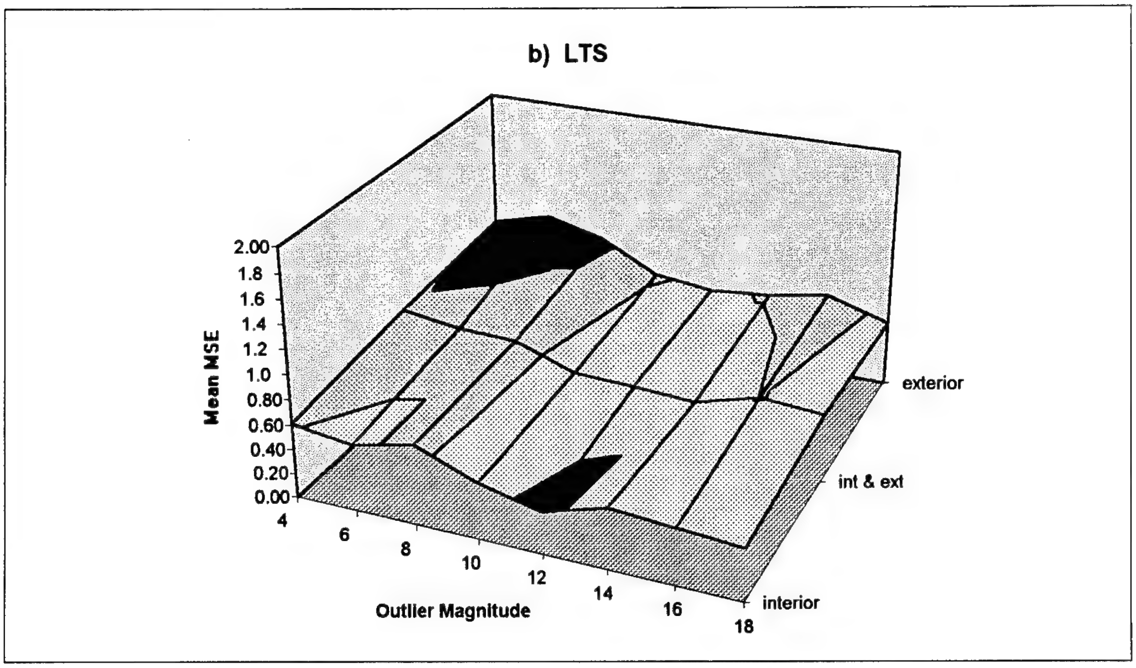


Figure 3.5. Pilot 1 Robust Regression Technique Overall Performance a) LMS and b) LTS

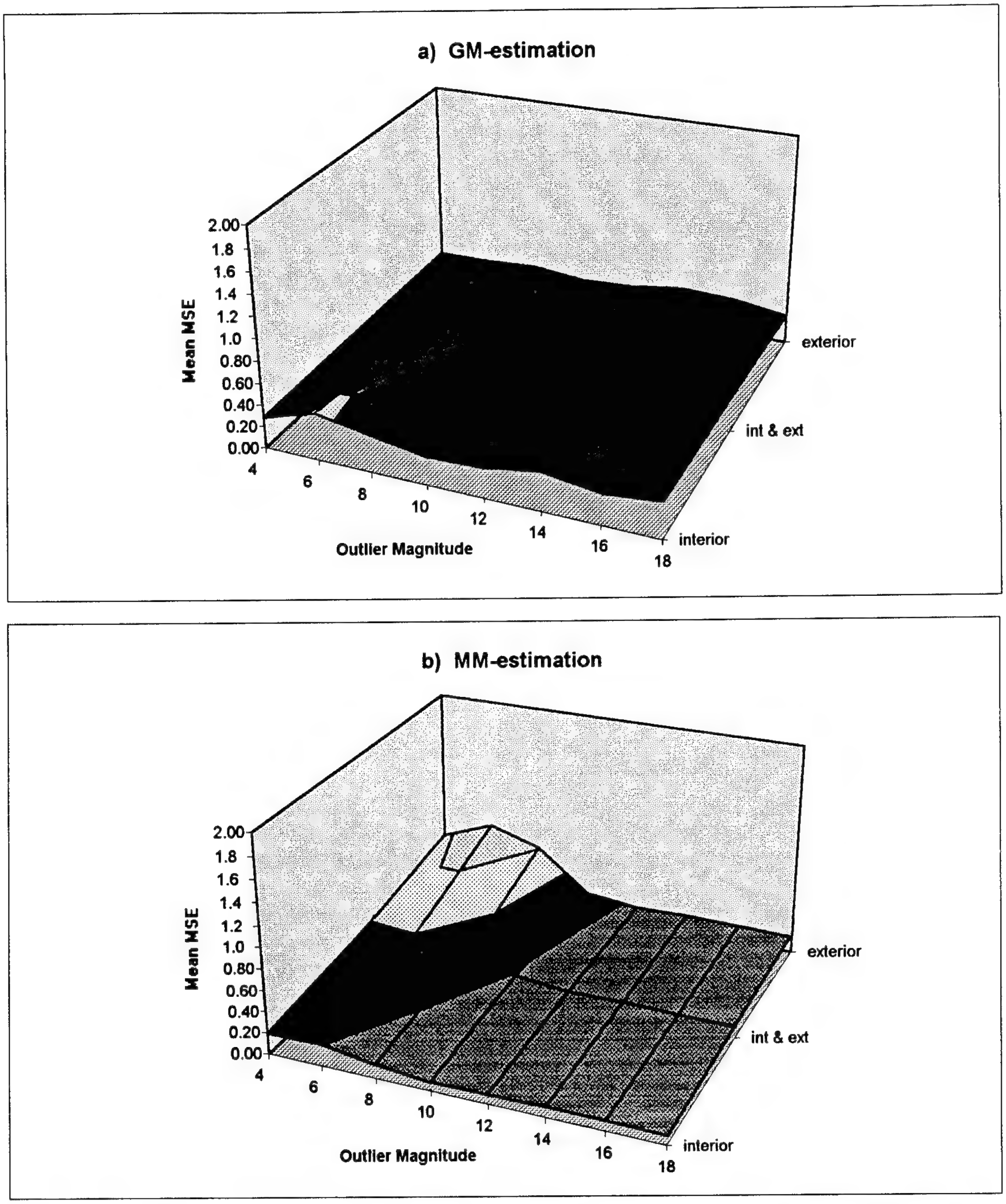


Figure 3.6. Pilot 1 Robust Regression Technique Overall Performance a) GM-estimation and b) MM-estimation

3.3.3 Observations on Outlier Magnitude

The 4σ case is fairly close to a nonoutlier or clean data situation. This configuration becomes a relatively good test of efficiency relative to least squares. As the outlier magnitudes increase, AMSEE values in general tend to improve to a certain point and then level off. This behavior is true for all the robust techniques and the level off point is similar across methods as well. Above 10σ, the AMSEE values level off. There is some change among the list of top performers as outlier magnitudes increase to 10σ, but their positions tend to hold steady after that point.

3.4 The Second Experiment (Pilot 2)

The purpose of the second study is to determine the effects that variable outlier magnitudes and locations have on robust regression technique performance. Specifically, we are interested in determining whether robust regression technique performance is affected by three factors:

1. Constant versus variable distance leverage points defining the subset of axial points
2. Constant versus variable magnitude of the errors for the outlying observations
3. Location of the outliers in the interior **X**-space region versus the exterior **X**-space region

The first two factors are changes to the study in Pilot 1, and the focus of the study in Pilot 2. Because the previous study involved using outlier errors and leverage distances of fixed length for all points, this second study considers the impact of varying these quantities among the six outlying points and eight leverage points in the data. The same regression techniques are in Pilot 2 so that comparisons can be made across pilot studies.

3.4.1 Dataset Development and Description of Regression Techniques

The design matrices for this study are nearly the same as those in Pilot 1, with one exception. The axial point distances are now a factor, so their values change across treatment combinations (see Table 3.6). Otherwise, the same model dimension, percentage of outliers, number of axial points, and signal-to-noise ratio is used.

Table 3.6. Fixed Factors for Pilot 2 Experiment Design

Number of variables	6
Sample Size	Design $2^{6-1} = 32 + 8$ axial = 40
Outlier Percentage and Number	15%, 6
Leverage (Axial) Points and Distance	8, Constant and Variable (see factor levels)
Coefficient Magnitude	4
Noise Level, Signal to Noise Ratio	1, 96 : 1

3.4.2 Experiment Design

In order to properly analyze the impact of outlier dynamics on the performance of various robust techniques, a three factor two-level design is used to capture the main effects and any possible interactions associated with the factors. This 2^3 factorial design consists of eight design points. The performance measures used in this study, the AMSEE and AMSIR, are the same measures used in Pilot 1.

The factor levels consist of: 1) constant versus variable outlier magnitudes measured by the size of the errors, 2) constant versus variable leverage distances measure by the axial points in the **X**-matrix, and 3) placement of the outlier errors in either the interior of the design space or in the high leverage points, resulting in exterior **X**-space outliers.

The outlier magnitudes selected in this study are based on the findings of Pilot 1. The results indicated that the performance of the techniques tended to improve as the outlier magnitudes increased to 10σ and then leveled off thereafter. The minimum magnitude such that the outliers were considered significant was about 6σ. Robust techniques with low efficiency did not perform as well as least squares for outlier magnitudes lower than 6σ. Based on these observations, the constant magnitude value was set at 10σ, while 6σ, 10σ, and 14σ were used for the variable magnitudes. Each dataset contains six outliers so the outlier magnitudes in the nonconstant scenario are 6σ, 6σ, 10σ, 10σ, 14σ, and 14σ.

The leverage distances used in Pilot 2 are larger than the 9.5 value used in Pilot 1 because it was originally believed that the 9.5 distance did not generate sufficient leverage. Errors were found in the ROBETH MVE robust distances algorithm and the necessary modifications were made. Subsequent analysis with the correct algorithm indicates that the 9.5 distance is indeed a high leverage position. However, the distances used in Pilot 2 are larger, which only means that the high leverage points have even more leverage. The factors and factor levels used in Pilot 2 are listed in Table 3.7. The outlier location factor is changed from a three-level factor in Pilot 1 to a two-level factor in this study for two reasons. First, the findings from Pilot 1 indicate that the factor level of both interior and exterior **X**-space outliers is simply a compromise between the two other levels and does not provide substantial additional information. Second, by moving to a two-level factor, the overall experiment design can be simplified and interpretation of the model terms is improved. The resulting design is a 2^3 full factorial in eight design runs.

Table 3.7. Pilot 2 Experiment Factors and Factor Levels

Factor	Levels
Outlier Magnitude	Constant (10σ), Variable (6σ, 10σ, 14σ)
Leverage Distance	Constant (20), Variable (16, 20, 24, 28)
Outlier Location	Interior **X**-space, Exterior **X**-space

3.4.3 Results of the Experimental Design

Based on the factors and factor levels described, separate analysis of variance models were developed for each technique. The response in each case is the technique's AMSEE. The 2^3 factorial design provides for independent estimates of the main effects and two-factor interactions, leaving the single three-factor interaction for the internal estimate of error. The model terms for each technique are estimated and the significant terms are selected by analyzing the normal probability effects plots. The terms identified for each model are shown in Table 3.8.

Table 3.8. Factor Levels Significant in Increasing the Technique's AMSEE

	Regression Technique						
Factor	LS	M	LAV	LMS	LTS	GM-2S	MM
Fixed or Varied Error	Varied						
Fixed or Varied Leverage				Fixed			
Interior or Exterior **X**-space Outliers	Exterior	Exterior	Exterior	Exterior	Exterior	Exterior	Exterior

The most significant overall factor is clearly the location of the outliers in the design space. This term describes the majority of the variability in the response, regardless of the technique measured. This result is not surprising knowing the high degree of leverage these axial points place on the model. The other factors are significant in only one model each. None of the two-factor interaction terms are significant.

Plots of some of the techniques' AMSEE values for each design point clearly reinforces the ANOVA findings (Figure 3.7).

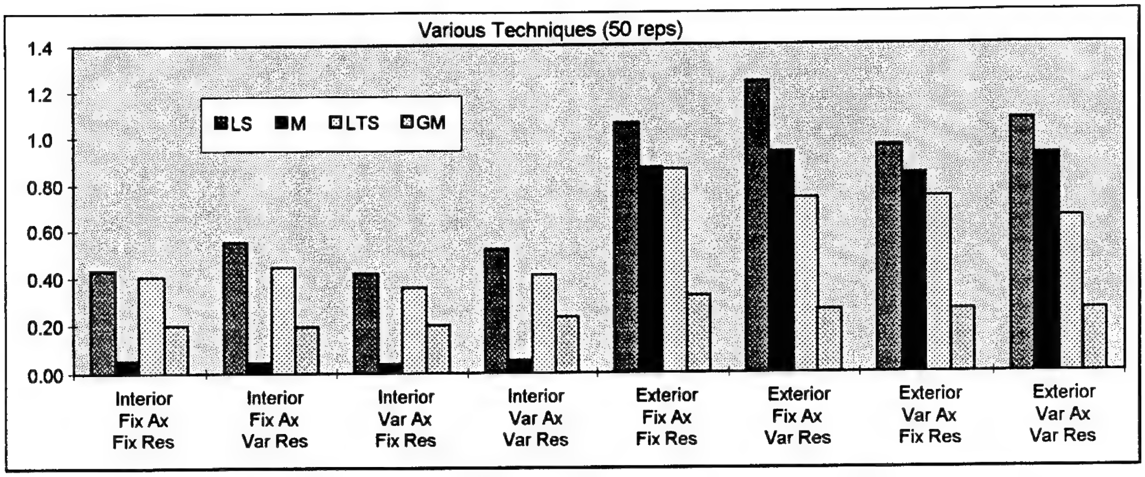


Figure 3.7. Pilot 2 Selected Robust Regression Technique Performances

Only the GM-estimation technique is somewhat unaffected by moving the outliers to the high leverage points. All of the other techniques show drastic relative increases in AMSEE. Although the MM-estimation AMSEE for exterior X-space outliers is about four times its interior X-space AMSEE, its performance relative to least squares and the other robust techniques is still very good (see Table 3.10). It is only outperformed by GM-estimation.

3.4.4 General Comments

The variance of the MSEE values differs across the various design points and various techniques. In some cases, a few poor fits caused fairly significant increases in the AMSEE. It will be interesting to see if a larger number of replicates (perhaps 50 instead of 30) lessen the impact of these poor fits.

The Pilot 2 AMSEE values are smaller (especially for LS) than the AMSEE values in Pilot 1. The smaller AMSEE values are caused by a decrease in the degree of estimation challenge to the robust techniques. The challenge is reduced when the leverage is increased on the axial points, while holding constant the magnitudes of the outlier errors. The larger axial distances cause these high leverage points to have even more pull on the parameter estimates. Because these leverage points are farther out in the X-space region, the outliers need to be proportionately farther away from the true line in order to have the same impact as the original outliers.

3.5 Additional Design Runs to Address Specific Issues

The general comments led to some questions needing answers before moving on to additional outlier studies. These questions are addressed by developing and running some sub-pilot studies consisting of a few to several runs per study. The studies will be discussed by first stating the issue to be addressed and then describing the design. The results are discussed in the following section.

Pilot 2a- The issue is whether fixed outlier magnitude datasets result in different AMSEE values than variable outlier magnitude datasets. This sub-pilot study is needed because the leverage distances are different between the two pilot studies and direct comparisons are not available. A two-run design is developed. The axial distances are reduced back to the Pilot 1 values and rerun. One treatment combination with fixed error magnitude, fixed axial distance, and exterior **X**-space, closely mimics the Pilot 1 treatment combination with a error of 10σ, and exterior **X**-space. The second design point (variable error magnitude, fixed axial distance, and exterior **X**-space) was run to determine whether varying error magnitude has an impact on AMSEE.

Pilot 2b- The issue is whether or not the variability of the AMSEE with 30 replicates is sufficiently small. Will the variability be reduced by running 50 replicates? To answer this question the initial Pilot 2 design is rerun with 50 replicates per design point so that proper comparisons can be made to the 30 replicate experiment.

Pilot 2c- The intent of this modification is to determine the technique performance impact of combining larger error values with the larger axial distances. The 50-replicate experiment in Pilot 2b is run with error magnitude changes. The errors for the fixed magnitude case are changed from 10 to 22, and for the variable magnitude case are changed from (6, 10, 14) to (18, 22, 26).

3.5.1 Sub-Pilot Study Results

Pilot 2a- Thirty replicates of the two design points described above are run and the AMSEE values are computed for each technique. The results compare favorably to the Pilot 1 design with 10σ errors in the high leverage or **X**-space location. The results for least squares is shown below in Table 3.9.

Table 3.9. Pilot 2a AMSEE Values for Fixed versus Variable Errors at Different Leverage Distances

Design	Pilot 2 (axial=20)	Pilot 2a (axial=9.5)	Pilot 1 (axial=9.5)
Fixed error, **X**-space	1.03	2.68	2.98
Varied error, **X**-space	1.22	2.95	

The second design point is run to determine whether a mix of error magnitudes affects technique performance. The results indicate that, for least squares, a small increase in AMSEE is observed. Recall however that least squares is the only technique affected by this factor (see Table 3.8).

Pilot 2b- To study the accuracy of estimation of location and dispersion for the mean square error of estimation (MSEE), box plots of the 30 replicate and 50 replicate runs are compared for the design point with fixed error, varied leverage, and exterior **X**-space outliers. There appears to be a slight decrease in the size of the 25th to 75th percentile box for most techniques. The variance of AMSEE is decreased mostly by the increase in the number of replicates ($V(\text{AMSEE}) = \hat{\sigma}^2/n$). The mean MSEE does not change much by increasing the number of replicates, indicating stability of the 30 replicate estimates.

The AMSEE values for each technique/design point combination at 50 replications produces results consistent with the 30 replication runs. The 50 replication runs do not result in AMSEE values that are significantly larger or smaller than the 30 replication runs. Table 3.10. shows the differences in AMSEE values between the 30 and 50 replicate runs. Due to the variability in error random variate samples between replicates, absolute AMSEE values will differ in run comparisons. It is more important to evaluate relative technique ranks for each treatment combination and identify differences. Table 3.10 also shows the relative rank differences between the 30 and 50 replicate runs.

Pilot 2c- The results of increasing error magnitude to accommodate the larger axial distances are not surprising based on the findings in Pilot 2a. Higher AMSEE values are observed for least squares and also for some of the robust techniques that do not perform very well with exterior **X**-space outliers. Comparison of the larger error runs with the smaller error runs, along with some comments regarding technique performance, are discussed in the following section.

Table 3.10. Pilot 2 Regression Technique Performance on Variable Outlier Magnitude and Axial Distance - 30 Replications versus 50 replications

AMSEE			30 Replications						
Vary Outlier Magnitude	Vary Axial Distance	Outlier Location	LS	M	LAV	LMS	LTS	GM	MM
No	No	interior	0.49	0.05	0.10	0.51	0.55	0.24	0.12
Yes	No	interior	0.58	0.06	0.10	0.46	0.43	0.26	0.13
No	Yes	interior	0.34	0.06	0.10	0.37	0.56	0.22	0.13
Yes	Yes	interior	0.54	0.04	0.07	0.30	0.41	0.17	0.11
No	No	exterior	1.03	0.91	0.98	0.86	0.88	0.28	0.48
Yes	No	exterior	1.22	0.92	0.99	0.96	0.77	0.27	0.46
No	Yes	exterior	0.94	0.87	0.93	0.82	0.80	0.29	0.53
Yes	Yes	exterior	1.17	1.07	1.15	0.78	0.85	0.28	0.47
			50 Replications						
Vary Outlier Magnitude	Vary Axial Distance	Outlier Location	LS	M	LAV	LMS	LTS	GM	MM
No	No	interior	0.43	0.05	0.07	0.35	0.41	0.20	0.11
Yes	No	interior	0.56	0.04	0.06	0.49	0.45	0.20	0.10
No	Yes	interior	0.42	0.04	0.07	0.33	0.36	0.20	0.10
Yes	Yes	interior	0.52	0.05	0.07	0.39	0.41	0.24	0.11
No	No	exterior	1.06	0.87	0.97	0.93	0.87	0.32	0.62
Yes	No	exterior	1.24	0.94	1.10	0.72	0.74	0.26	0.46
No	Yes	exterior	0.97	0.85	0.98	0.80	0.74	0.27	0.45
Yes	Yes	exterior	1.07	0.93	1.02	0.81	0.66	0.27	0.43
			Difference in Technique Ranks						
Vary Outlier Magnitude	Vary Axial Distance	Outlier Location		M	LAV	LMS	LTS	GM	MM
No	No	interior		0	0	0	0	0	0
Yes	No	interior		0	0	0	0	0	0
No	Yes	interior		0	0	0	0	0	0
Yes	Yes	interior		0	0	0	0	0	0
No	No	exterior		1	0	-2	1	0	0
Yes	No	exterior		-1	0	2	-1	0	0
No	Yes	exterior		0	0	0	0	0	0
Yes	Yes	exterior		0	0	-1	1	0	0

3.5.2 Technique Performance in the Smaller Error Study

The performance of the high efficiency group of estimators, *M* and LAV estimation, are very similar in both absolute and relative terms. The same comparison is true of the high breakdown estimators, LMS and LTS. Regarding the outlier factors used in these studies, the two techniques in each group have AMSEEs that are nearly identical. Effective candidate technique reductions involves eliminating one technique from each group in future tests. Although the multi-stage estimators, GM and MM-estimation also perform somewhat similarly, enough differences exist to warrant further testing of both techniques. An additional reason for examining these techniques further is that they perform better than any of the other techniques to this point.

Both *M* and LAV estimation achieve near perfect estimation of the model parameters when outliers are located in the interior of the **X**-space region. AMSEE values range from 0.04 to 0.07, compared with AMSEE values of 0.50 for LS. These high efficiency techniques also perform better than any other robust methods for interior **X**-space outliers. However, they perform the worst of the robust methods when outliers are placed in the exterior **X**-space region. In fact, they did not perform any better than LS. LMS and LTS are average performing robust techniques in these studies. They perform slightly better than LS in all cases (e.g. AMSEE values of 0.40 for interior **X**-space compared with 0.50 for LS).

As previously mentioned, GM and MM perform the best. For interior **X**-space outliers, MM was clearly better with AMSEE values of 0.10 compared to GM values of 0.20. In the exterior **X**-space scenarios, where the bounded influence techniques should perform well, the GM technique clearly outperformed all other robust methods, posting the best AMSEE of 0.30 compared to the second best numbers (MM estimation) of 0.50.

3.5.3 Technique Performance in the Larger Error Study

By increasing the values of the outlier errors, the least squares AMSEE values increase substantially and the robust estimations are challenged more thoroughly. The LS AMSEE values of about 2.0 for interior **X**-space outliers and 5.0 for exterior **X**-space outliers, are four to five times higher than the values in the smaller error study (Table 3.11).

Table 3.11. Performance of Least Squares on Small versus Large Error Designs (AMSEE)

Vary Error Magnitude	Vary Axial Distance	Outlier Location	Small Errors	Large Errors
No	No	interior	0.43	2.13
Yes	No	interior	0.56	1.85
No	Yes	interior	0.42	2.29
Yes	Yes	interior	0.52	1.87
No	No	exterior	1.06	5.10
Yes	No	exterior	1.24	5.35
No	Yes	exterior	0.97	4.49
Yes	Yes	exterior	1.07	4.88

In terms of technique performance, robust methods from the same class behaved in similar fashions. The estimation errors for *M* and LAV techniques are nearly indistinguishable as were the LMS and LTS estimation errors. *M* and LAV estimations are again impressive under interior **X**-space outliers but miserable when exterior **X**-space outliers are present. The LMS and LTS AMSEE values are almost unchanged from the small error study, so when compared to the large error LS values, these techniques perform respectably.

The GM and MM again performed the best of all robust techniques tested. Like LMS and LTS, the GM estimation error values were very similar to the GM estimation error values from the small error study (0.20 for interior **X**-space and 0.30 for **X**-space). This finding suggests that the technique downweights residuals to a certain point but will not give them zero weight. This particular version of GM-estimation uses the Huber ψ-function which is a monotone function that assigns the weight to be a ratio of the tuning constant to the scaled residual. Improved estimation with larger errors may be obtained by using one of the popular redescending ψ-functions such as Tukey's biweight. The MM-estimation error values for the interior **X**-space outliers were also unchanged from the small error study, but the AMSEE values under the exterior **X**-space outliers decreased from 0.50 to about 0.23. It appears that the larger error points have been further downweighted, perhaps to zero, resulting in an improved estimation relative to the true equation. The results of GM versus MM for small and large errors are plotted in Figure 3.8. Increasing the magnitudes of the errors favors MM-estimation in this comparison.

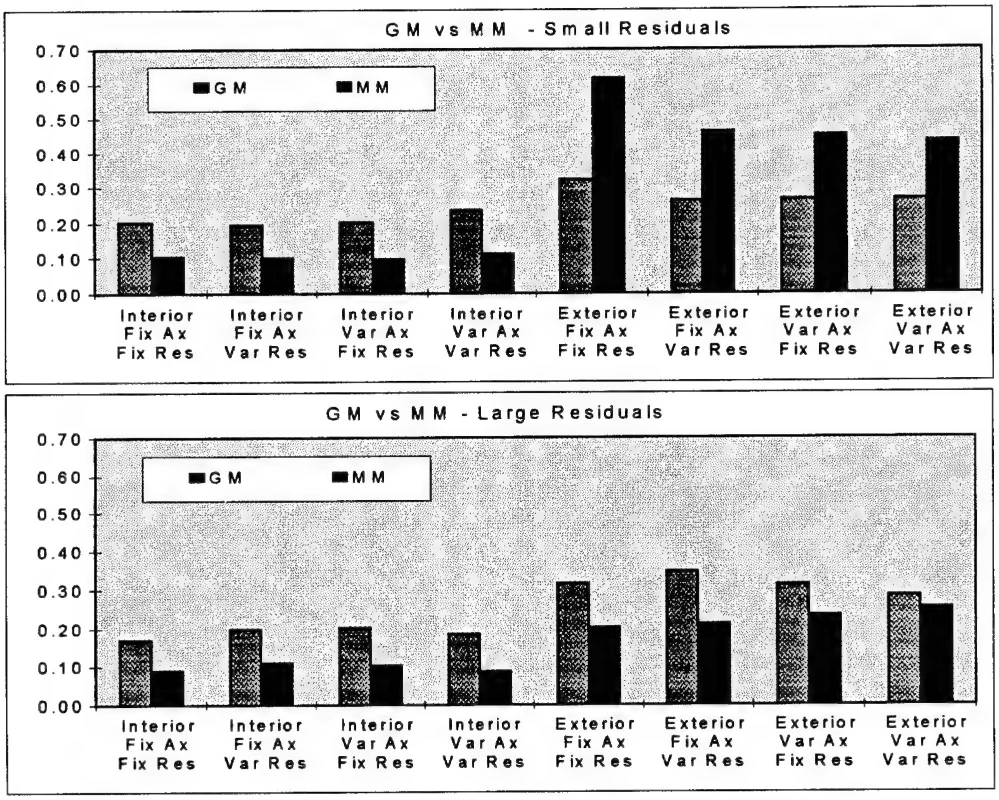


Figure 3.8. Pilot 2 GM-estimation versus MM-estimation on Small and Large Error Outliers

3.6 Conclusions

The major findings of these two studies are presented in the following list. Some of the findings are notes worth recalling in future outlier designs. Other findings relate to the performance and possible enhancement of robust techniques, also for subsequent study.

- Regarding dataset design, increases in axial distances should be accompanied by respective increases in error magnitude
- Increasing the number of replications may improve accuracy of estimation of the AMSEE, especially in terms of reducing the impact of poor fits on the final estimate
- Varying axial distance or varying error magnitude has only a slight, insignificant impact on robust technique performance
- A comparison of robust distances between the fixed leverage and varied leverage designs indicates that a range of robust distances exists for both the fixed and varied case. Also, because complete search using the MVE approach is time and cost prohibitive (18 million subsets taking 7 hours on a 486DX2-66), widely different answers are obtained with random subset searches
- Considering the various outlier magnitude and location factors used in the first two studies (which does not necessarily include high breakdown point because only 15% outliers are considered), two techniques outperform all others: GM and MM-estimation. Enhancements to either or both techniques should be considered in a search for the best overall technique

- A possible enhancement to the GM technique is to consider using a redescending ψ-function to drive the large error, high leverage points to a zero weight. Even a small nonzero weight of large outliers can result in less than favorable estimates

Some of these findings are crucial in terms of designing proper outlier configurations. The relationship that exists between outlier magnitude and degree of leverage may appear intuitive in hindsight, but was certainly not anticipated in the original designs. The screening of various robust techniques has been invaluable information for deciding which types of methods are superior across a variety of situations, and for determining possible enhancements that can be incorporated into the design of new robust techniques. Much of the insight gained has improved the quality of the work accomplished in future chapters.

Chapter 4

The Development and Evaluation of Generalized *M*-estimation Techniques

4.1 Introduction

Robust regression techniques are designed to build useful regression models of data containing outliers, often referred to as nonnormal data. Three properties characterize robust techniques in terms of their ability to fit models to nonnormal data. The first two properties, high breakdown and bounded influence refer to the capability of a technique to fit a model to data containing outliers of an increasing quantity and distant **X**-space location. The third property, high efficiency, compares robust techniques to least squares estimation when the errors are normally distributed and no outliers are present.

The classic technique of least squares estimation has weaknesses when outliers occur in the data. One of these weaknesses is breakdown, a phenomena concerning the outlier density of the data. The breakdown is the percent of outliers allowed in the data before the estimator fails to give adequate information regarding the relation between the regressors and the response. The highest possible breakdown is 50%, so that up to half of the data can be spurious and a technique should be able to fit a model to the "good" half of the data. For least squares the breakdown is 0%, meaning that only one sufficiently large outlier can significantly alter estimates in least squares.

Another shortcoming of least squares is that points outlying in the regressor, or **X**-space, have substantial influence on the model coefficient estimates. The influence of these points is not bounded, meaning that the farther the points are located in **X**-space relative to the centroid of the **X**-space, the more influence they have on the estimation of the model parameters. If these influential points are not "in-line" with the other observations, the least squares estimation is pulled in the direction of the influential points. A desirable property of robust estimators is for their influence function to be bounded in the **X**-space so that the points located far from the **X**-space centroid do not have significantly higher influence over the other points.

Robust estimators are also desirable if they perform nearly as well as least squares when the errors are normally distributed. This characteristic, called efficiency, is measured by ratio of the least squares mean square error to the robust technique mean square error. Efficiencies close to one are desirable. Techniques that perform well in the presence of normal *and* nonnormal conditions would satisfy the true definition of robust, meaning they can be appropriately used in any data configuration.

Among the types of robust regression techniques available, generalized *M*-estimation (GM-estimation), also called bounded influence estimation, provides the ideal framework for developing an estimator containing all three of the desired robust properties of high breakdown, bounded influence and high efficiency. Obtaining these properties would enable a technique to: 1) accurately estimate model parameters in the presence of a high percentage of outliers (up to 50%), 2) exert control over the **X**-space outliers by adequately downweighting the high leverage points with large standardized errors, and 3) provide estimates similar to least squares under normal error conditions.

The purpose of this chapter is to evaluate existing GM-estimation techniques and to develop GM technique alternatives that may improve estimation across a variety of outlier and nonoutlier scenarios. Three existing methods are selected that represent a variety of possible approaches and offer the most promising performance characteristics. A number of potentially significant enhancements to these methods are proposed and evaluated in a combined experiment. Several pilot studies are performed with the experiment that aid in the development of improved alternatives. The resulting alternatives are compared using performance measures, including mean square error of estimation, technique relative ranking, and mean square inefficiency ratios.

4.2 Background

The general approach for GM-estimation, detailed in Chapter 2, is to enhance the approach of *M*-estimation which involves finding the minimum of a function of the residuals that increases less rapidly than the least squares squared function. Recall that the objective of *M*-estimation is to minimize

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \rho\left(\frac{e_i}{s}\right) = \min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \rho\left(\frac{y_i - \mathbf{x}_i'\hat{\boldsymbol{\beta}}}{s}\right) \tag{4.1}$$

Taking the partial derivatives of the objective with respect to β, and defining $\psi = \rho'$, the system of equations can be written

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \psi\left(\frac{e_i}{s}\right) = \min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \psi\left(\frac{y_i - \mathbf{x}_i'\hat{\boldsymbol{\beta}}}{s}\right)\mathbf{x}_i = \mathbf{0} \tag{4.2}$$

The *M*-estimator bounds the influence of the observations in the interior **X**-space (influence of the residual), but does not bound the influence of outliers in the regressor or **X**-space (influence of position). GM-estimators (of the Schweppe type) bound the influence of position in addition to the influence of residuals by weighting the *M*-estimation system of equations

$$\sum_{i=1}^{n} \pi_i \psi\left(\frac{y_i - \mathbf{x}_i'\hat{\beta}}{s\pi_i}\right)\mathbf{x}_i = \mathbf{0} \tag{4.3}$$

The weights $\pi_i = \pi(\mathbf{x}_i)$ are a function of the measure of the distance $\mathbf{x}_i$ from the center of the multivariate regressor space cloud.

Computation of GM-estimates require two stages of coefficient estimation, consisting of an initial estimate that provides a good starting point, followed by convergence to the final GM-estimate. Development of a technique requires that a number of decisions be made regarding choices of initial estimators, estimators of scale and leverage, type of π-weight and ψ-function, and type and amount of convergence necessary (Table 4.1).

Table 4.1. GM-Estimation Technique Characteristics

GM-Component	Comments
Bounded Influence Objective	The preferred type is that of Schweppe, who proposed an objective that theoretically downweights high leverage points only if the residual is large
Initial Estimate	The intent is to provide a good starting point. High breakdown estimators are typically used
Estimate of Scale	Several high breakdown choices are available including the MAD, the LMS estimate of scale, and the scale output of the initial estimate (from *S*-estimates). The scale estimate can be updated in final estimate iterations, but convergence is not necessarily assured
Estimate of Leverage	Different methods are available. A tradeoff exists between computational ease and the ability to handle clouds of multiple outliers
π-weights	Several different approaches are available corresponding to the type of leverage measure used. Some approaches require inlier/outlier cutoff values
ψ-function	*M*-estimate residual downweighting functions including Huber's *t*, Tukey's biweight and Ramsay's exponential
Tuning Constant (ψ-function)	Depends on the ψ-function and desired efficiency. Sometimes it is also a function of *n* and *p*
Convergence	Nonlinear convergence algorithms include, for example, Newton's method, and (IRLS). Another consideration is the number of iterations so that the initial estimate breakdown property is preserved

4.3 Published Techniques

Several GM-estimation methods consisting of various combinations of GM-components have been proposed. Three of these proposals will be implemented and tested against several datasets. In addition, variations of the three proposals will be developed in an effort to find the best performing technique. The published techniques consist of proposals by Walker (1984), Coakley and Hettmansberger (1993), and Marazzi (1993). These techniques are selected for different reasons but together comprise a diverse and encompassing set of alternatives. As the nickname implies, all GM techniques have bounded influence. Each proposal and all of the variations in this paper use the Schweppe objective function, which many prefer over the other popular GM objective proposed by Mallows. The Mallows objective tends to downweight high leverage residuals regardless of the size of the residual, while the Schweppe estimator theoretically only downweights high leverage points with large residuals (Krasker and Welsch 1982 and Hampel et al. 1986 p. 322, among others).

The GM-estimation approach of Walker (1984) is one of the more standard approaches in terms of its GM-components. Emphasis in the development of this estimator was placed on computational ease, high efficiency, and the ability to accurately estimate under the combined problem of outliers *and* multicollinearity. Walker's method for bounding the influence is based on an earlier suggestion of Welsch (1977) which constructs the *π-weights* using the hat matrix diagonals in such a fashion that the argument of the ψ-function is similar to the *DFFITS* influence diagnostic proposed by Belsley, Kuh, and Welsch (1980). The *DFFITS* diagnostic is designed to reveal the influence of the i^{th} observation on the fitted or predicted value. This diagnostic can be computed as a function of the hat diagonals, the residuals and an estimate of scale such that

$$\begin{aligned} DFFITS &= \left(\frac{h_{ii}}{1-h_{ii}}\right)^{1/2} \frac{e_i}{s_{(i)}(1-h_{ii})^{1/2}} \\ &= \left(\frac{h_{ii}{}^{1/2}}{1-h_{ii}}\right) \frac{e_i}{s_{(i)}} \end{aligned} \tag{4.4}$$

By equating the inverse of the first term above to the π-weight and using some robust estimate of scale for $s_{(i)}$, the DFFITS expression can be incorporated in the argument of the GM ψ-function so $\psi(e/\pi s) = \psi(DFFITS)$. The other GM-components of the Walker method help achieve the high efficiency and computational ease characteristics. The initial estimator used is least squares, and a non-iterated MAD is recommended as the estimate of scale. Convergence to the final estimate is obtained using fully iterated reweighted least squares. High breakdown was not a primary consideration in the development of this estimator primarily because interest in this property did not become prevalent in the research until after Walker's method was published.

The method of Coakley and Hettmansberger (1993) was proposed to be the first GM-estimator with high efficiency, high breakdown and bounded influence. The approach adopts some of the more respected techniques for the GM-components, and suggests a convergence approach that seeks to maintain high breakdown in the final estimate. The initial estimate consists of the high breakdown LTS estimator and uses the LMS scale estimate, which is also high breakdown. The robust distances using the MVE estimator (Rousseeuw and van Zomeren 1990), are used for estimates of leverage. Cutoff values for leverage outliers are determined using a χ^2 statistic with $\alpha \cong 0.025$ and p-1 degrees of freedom. For the actual π-weights, the authors propose using a ratio of the χ^2 cutoff value to the squared Robust Distances (bounded below by 1) as the π-weights.

$$\pi_i = \min\left[1, \left\{\frac{\chi^2_{0.025,p-1}}{RD^2}\right\}\right] \tag{4.5}$$

The primary drawback of this approach is the computational intensity involved to compute both the initial estimate and the estimates of leverage.

The proposal of Marazzi (1993) in many ways is a compromise between the Walker and Coakley-Hettmansberger methods. Marazzi proposes initial estimates and estimates of leverage that are robust, but not necessarily high breakdown. The benefits of their approach are the potential for higher efficiency and increased stability in the estimates. Instability is often associated with high breakdown estimates because they require random subsampling methods that generate approximations. These approximations can sometimes contain significant variability around the exact estimate. The robust method of Marazzi is computationally more efficient than the high breakdown methods. He proposes the most B-robust estimator of Hampel et al. (1986, p 318) which consists of an LAV estimator weighted by estimates of leverage. Weights proposed by Krasker and Welsch (1982), which are related to *M*-estimates of covariance, are used to generate the leverage distances in the **X**-space region. The π-weights are computed as the inverse of these distances. Their original proposal suggests using a fully iterated Newton's algorithm for convergence, but initial tests show little difference between Newton's method and IRLS. Thus, the IRLS method is used in performance experiments.

The GM-components for each of the proposals is shown in Table 4.2. A detailed discussion of the component techniques is provided in Chapter 2. A variety of components are used among the three proposals.

Table 4.2. Published Candidate Bounded Influence Regression Techniques

	Technique		
Component	**Walker (1984)**	**Coakley and Hettmansperger (1993)**	**Marazzi (1993)**
GM Objective	Schweppe	Schweppe	Schweppe
Initial Estimate	LS	LTS	Most B-Robust
Scale Estimate	MAD	$\hat{\sigma}_{LMS}$	MAD
Leverage Measure	h_{ii}	Robust Distance (based on MVE)	Krasker-Welsch weights (z)
π-weight Function	$h_{ii}^{1/2}$ / (1-h_{ii})	min(1, b/RD^2)	1/|z|
ψ-function	Huber	Huber	Huber
Tuning Constant	1.345·f(n,p)	1.345	1.05·$p^{1/2}$
Convergence Approach	Fully Iterated IRLS	One-Step Newton-Raphson	Fully Iterated IRLS
Properties			
High Efficiency	Yes	Yes	Yes
High Breakdown	No	Yes	No
Bounded Influence	Yes	Yes	Yes

4.4 Proposed Alternatives

The three proposals are used as baselines for developing alternative techniques that either employ different variations of the original components or adopt other promising components. The objectives of the proposal modifications are to:

- Increase the breakdown point of Walker and Marazzi methods (initial estimates and measures of leverage)
- Increase the efficiency of the high breakdown methods by using more efficient high breakdown initial estimates (such as *S*-estimators)

- Substitute computationally more stable high breakdown measures of leverage (such as the MCD) for the MVE estimates
- Introduce redescending (hard and soft) ψ-functions into proposals using monotone ψ-functions
- Use IRLS convergence if possible. This convergence technique can be used to develop estimators that can address the combined outlier/collinearity problem
- Modify the π-weight measures as necessary to prevent substantial downweighting of low residual observations
- If possible, decrease the computational requirements of the proposals without sacrificing performance
- Decrease the variability of the estimates by avoiding methods using random subsampling algorithms

Using these objectives, alternative candidate GM techniques are developed, evaluated and compared with the three published methods. In developing each alternative, preliminary estimation performance was measured by computing the MSEE of a six-variable dataset with high leverage points and outliers in the high leverage observations. Adjustments and decisions for further testing were made based on this series of initial tests. As a result of this screening process, the following observations are provided.

- The final IRLS weights ($w_i = \psi(e_i / \pi_i s) / (e_i / \pi_i s)$) of the Marazzi method indicate that severe downweighting occurs for non-outlying observations. This behavior is primarily caused by the magnitudes of the π-weight values used in the computation of the w_i. The Krasker-Welsch weights, which are measures of leverage distances, have significant magnitude even for the low leverage points. The π-weight values are equal to the inverse of the Krasker-Welsch weights, so large Krasker-Welsch weights mean small π-weight values. Small π-weights result in a) large arguments for the ψ-function, meaning small values for $\psi(z)$ and b) large denominators for w_i. These two conditions result in small weights, regardless of the residual magnitude or leverage degree. One possible remedy to this problem is to scale the large magnitude leverage distances by dividing each distance by the median distance in the dataset. Thus, the median distance has a scaled value of one, and the other scaled distances are just the ratio of the original distance to that median distance. This approach was implemented in alternatives MZ1, NP1, and NP2

- The one-step Newton convergence technique suggested by Coakley and Hettmansberger may indeed preserve the high breakdown aspect of the initial estimate, but initial tests with the six-variable dataset indicate that little change occurs in the coefficient estimates with one Newton step
- The LTS initial estimates are obtained using a random subsampling algorithm. This algorithm produces coefficient estimates that can vary considerably
- Initial results indicate that the type of ψ-function used does not significantly change the final coefficient estimates
- No alternative technique appears to produce a significantly improved estimator over any of the original proposals
- The most B-robust estimator (LAV estimator weighted by leverage) and *S*-estimators tend to provide good starting point estimates
- The rate of convergence for the fully iterated Newton's method versus the fully iterated IRLS method is similar and the final estimates are nearly identical

The alternative techniques are assigned alpha-numeric labels that indicate the originating proposal and number of the variation. The alternatives labeled NP are techniques that differ substantially from the original proposals. Including the three original proposals, a total of eleven robust GM techniques are presented for comparison. Two alternatives consisting of components of different proposals plus some suggestions provided in this study are labeled NP (New Proposal) followed by a number variation. Table 4.3 lists the components of the alternative techniques and highlights (shaded areas) the components that are different from the original proposal.

Table 4.3. Alternative GM-Regression Techniques

GM-Component	**WA1**	**CH1**	**CH2**	**CH3**	**MZ1**	**MZ2**	**NP1**	**NP2**
GM Objective	Schweppe	Schweppe	Schweppe	Schweppe	Schweppe	Schweppe	Schweppe	Schweppe
Initial Estimate	***S*-estimator**	LTS	***S*-estimator**	***S*-estimator**	Most B-Robust	Most B-Robust	Most B-Robust	*S*-estimator
Scale Estimate	$\boldsymbol{\hat{\sigma}_{S-est}}$	$\hat{\sigma}_{LMS}$	$\boldsymbol{\hat{\sigma}_{S-est}}$	$\boldsymbol{\hat{\sigma}_{S-est}}$	MAD	MAD	MAD	$\hat{\sigma}_{S-est}$
Leverage Measure	h_{ii}	RD - MVE	RD - MVE	RD - MVE	KW weights.	**RD - MVE**	KW weights.	KW weights.
π-Weight Function	$h_{ii}^{1/2}$ / (1-h_{ii})	min(1, b/RD^2)	min(1, b/RD^2)	min(1, b/RD^2)	**med|z| / |z|**	**min(1, b/RD^2)**	med |z| / |z|	med |z| / |z|
ψ-Function	Huber	Huber	Huber	**Biweight**	Huber	Huber	Huber	Huber
Tuning Constant	4.685·f(n,p)	1.345	1.345	1.345	$1.05 \cdot \sqrt{p}$	**1.345**	1.345	1.345
Convergence	Iterated IRLS	**Iterated IRLS**	**Iterated IRLS**	**Iterated IRLS**	Iterated IRLS	Iterated IRLS	Iterated IRLS	Iterated IRLS

Note: WA: Walker; CH: Coakley and Hettmansberger; MZ: Marazzi; NP: New Proposal

4.5 Technique Performance Experiment

The dataset screening previously mentioned was designed to gain some insight into possible improvements on the original proposals and to establish a direction to search for potential alternative techniques. The next step is to compare these alternatives among each other and with the original proposals by developing a series of datasets designed to challenge the capability of each technique to perform well under a variety of outlier conditions. The process involves developing a series of datasets that will expose possible weaknesses in the techniques, estimating the techniques' model parameters, analyzing their performances, and summarizing the results.

4.5.1 Developing the Datasets - Multiple Point Clouds

An approach similar to previous screening experiments is adopted using a Monte Carlo data generation scheme and model development so that coefficient estimates can be compared to the true coefficient values. The same basic factorial and fractional factorial designs augmented by high leverage points is used so that collinearity between the model regressors is not present and does not confound the outlier influence issue. The purpose of dataset development is to create a reasonable number of scenarios to 1) test the capability of various measures of **X**-space leverage to locate high leverage points and 2) test the properties of high breakdown, high efficiency, and bounded influence.

Efforts have focused not only on the problem of identifying single points, but also on the more difficult multiple outlying points in a multivariate region. Like robust linear modeling, many techniques have been proposed and significant progress has been made, but no approach lacks vulnerability. The three most common measures of **X**-space location and dispersion are imbedded in the GM approaches being considered. These techniques are the hat matrix diagonals, the *M*-estimates of covariance (using the Krasker-Welsch proposal), and the robust distances using the MVE. One of the more difficult outlier scenarios for each of these approaches involves identifying subset(s) of multiple points forming a cloud that is separated from the majority of the points in the set. The distance and relative position of the cloud from the rest of the data can often obscure it from detection as high leverage by some or all of these techniques. One of the challenging aspects of these clouds is that they often are composed of points that are not outliers individually, but taken together can be influential. An attempt is made to design a dataset containing a high leverage, multiple point cloud that escapes detection by one or more of the three leverage estimation techniques.

Each dataset consists of a two-variable design of 12 interior points (3 replicates of a 2^2 factorial design), and 4 high leverage points in a cloud located away from the 12-point cube. The total design space consists of 16 points, 25% of which are **X**-space outliers. The purpose of the study is to identify vulnerabilities in the leverage estimation methods. The goal is to find locations for the 4-point cloud that are significantly far away from the cube, but not easily detected by one or more of the location/dispersion techniques. Hawkins, Bradu and Kass (1984) developed a now famous data set consisting of 75 observations containing 14 high leverage points. In developing this dataset, Hawkins, Bradu and Kass use the multiple point cloud approach to locate all 14 leverage points near each other and away from the other 61 observations. Using this data, Rousseeuw and van Zomeren (1990) show that the hat matrix diagonals only identify one of the 14 high leverage points as an **X**-space outlier. They also show that their robust distances method using MVE correctly identifies all 14 points as **X**-space outliers. The multiple point cloud configuration clearly identifies a weakness in the hat diagonal's approach to detect high leverage points. The shortcoming of the hat diagonal's metric here is that the computation of center of mass, which is used as the origin for measuring leverage distances, is influenced by the outlying observations. The centroid is moved in the direction of the outliers, so the resulting distances of the outliers are not large relative to the inlier distances.

These types of dataset configurations also reveal vulnerabilities in the Krasker-Welsch weights and MVE robust distances. Krasker-Welsch weights, like the hat diagonals, are not high breakdown estimators and sometimes fail for the same reason. If the percentage of high leverage points stays below $1/p$, as in the Hawkins-Bradu-Kass (H-B-K) dataset, the Krasker-Welsch method works well. All fourteen of the outliers are clearly identified using the KW method. However, if the percentage of high leverage points exceeds $1/p$, the high leverage point distances are not significantly higher than the interior point distances. For example, if the last 30 observations of the H-B-K dataset are deleted, resulting in a dataset with over 30% high leverage points, the KW method does not perform well. The centroid is influenced sufficiently by the outliers so that the resulting Krasker-Welsch high leverage distances are not large relative to the inlier distances.

The robust distances based on the MVE estimator have a breakdown of 50% if the subsample sizes are chosen correctly. The estimator used in all tests was a 50% breakdown estimator. However, using datasets with only 25% high leverage points, the technique failed to detect some outlying cloud locations. Remembering that the 50% breakdown MVE estimator is designed to identify the minimum volume ellipsoid covering just over half of the data, it is possible to locate the high leverage points far away from the cube but within the MVE. Placing the cloud in line with points in the cube results in an minimum volume ellipse that is elongated and narrow, covering the high leverage points. Figure 4.1 illustrates this scenario.

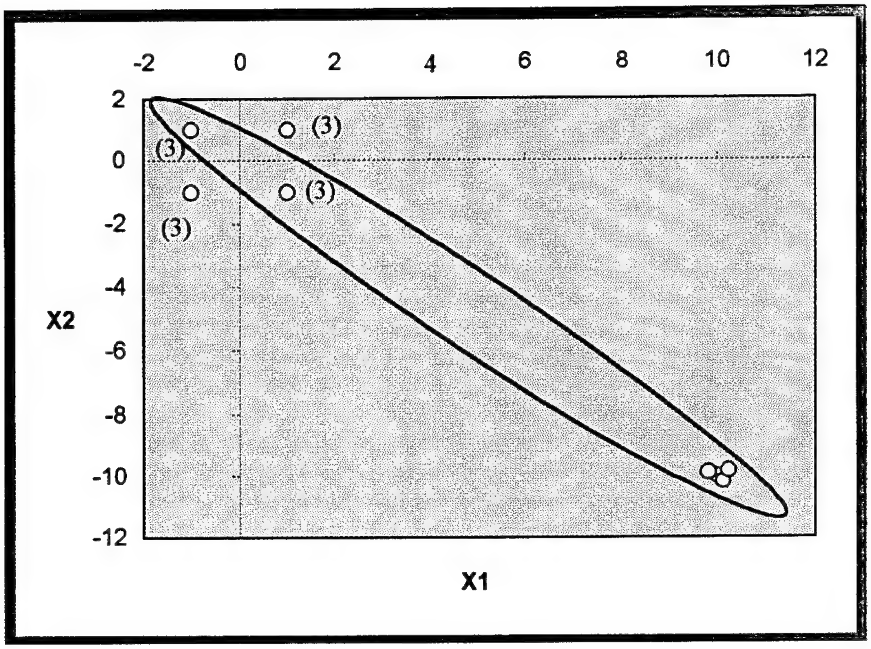


Figure 4.1. Minimum Volume Ellipsoid Covering Outlying Point Cloud

Not only does the MVE technique "mask", or fail to correctly identify the discrepant cloud, but it also "swamps" the off-diagonal cube elements, meaning that inliers are incorrectly specified as outliers. In this example, these off-diagonal cube points have robust squared distances that are nearly fifteen times the cutoff value suggested by Rousseeuw and van Zomeren (1991).

Interestingly, the MVE is drawn around this cloud located a large distance from the cube along the diagonal only if there is some variability in the points in the cloud. If the four points are replicates positioned on the diagonal (10, -10), the MVE correctly identifies the cube points as inliers. It appears that the ellipsoid requires some relative spread in each dimension.

Another cloud location using this configuration reveals a different dynamic of the MVE concept. Consider placing the cloud (not all points identical) along one of the axes and steadily increasing its distance from the cube. Figure 4.2 shows how the MVE changes as the cloud moves away from the cube. Due to the alignment of the cloud relative to subset points of the cube, the MVE technique does not identify the cloud as outliers until those points are moved a considerable distance from the cube points, indicating perhaps that the power of this technique may not be high for this type of scenario. Simonoff (1991) performed Monte Carlo simulations on data using MVE robust distances. High leverage points were identified as observations with distances exceeding the χ^2 (α=0.025, p-1) cutoff. Simonoff found that, on the average, the robust distances technique leads to 5 out of 20 cases being identified or swamped in clean data.

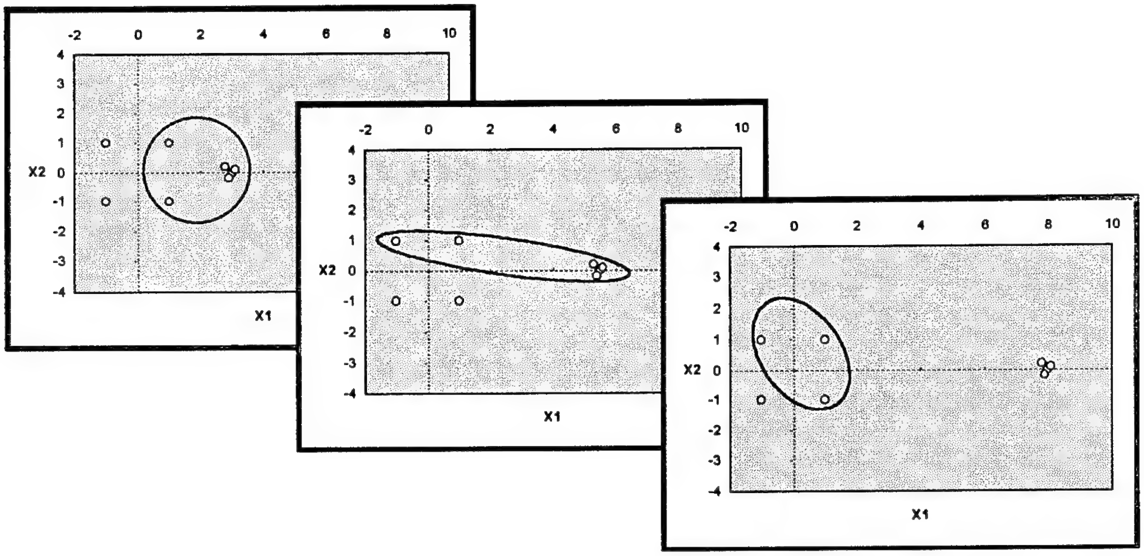


Figure 4.2. Minimum Volume Ellipsoid Dynamics as Cloud Moves Along X1 Axis

Other techniques, such as the Krasker-Welsch weights and the hat diagonals, also have trouble identifying the outlying cloud as significantly far away from the rest of the points. Because the cloud causes the centroid to move away from the majority of points represented by the cube, the cloud points have only slightly larger distances than the cube points furthest from the new centroid.

Both Krasker-Welsch weights and the hat diagonals represent improvements over MVE distances because huge distances are not assigned to actual inliers (off-diagonal cube points) as they are in the MVE approach. Clearly, each of the diagnostics has weaknesses indicating the dangerous nature of the multiple point cloud problem.

A third cloud location, not in-line with the cube points (at (10, -7) for example), favors the MVE approach over the other two approaches. The Krasker-Welsch weights and the hat diagonals perform as described in the previous case, giving similar weights to the cloud points and some of the cube corner points. The MVE approach identifies the cloud correctly and gives all the cube corner points similarly small distances.

The performance of these techniques obviously depends on the situation. Relative to multiple point clouds, the MVE approach either succeeds or fails miserably by assigning high distances to interior **X**-space points. The Krasker-Welsch weights and hat diagonals do not perform well in general but also do not make huge mistakes by incorrectly assigning large weights to interior points.

All three of these cloud point datasets generate a degree of collinearity among the two independent variables, but not to a point that the variability of parameter estimates are significantly increased. The diagnostics used to measure the collinearity between the X variables is the variance inflation factor (VIF). VIF values larger than 10 are regarded as cause for moderate concern. Of the three datasets analyzed, none of the VIF values exceed 6.0. Although a slight dependency exists, it is not large enough to significantly alter the accuracy of the parameter estimates.

4.5.2 Selected Datasets

The intent of this study is to determine the performance of a number of GM techniques in terms of their ability to fit models to the "good" observations from a set of data. Their performance relative to contaminated data will be measured by comparing the technique coefficient estimates to the true model coefficients. Desirable properties are tested by appropriate quantity and location of outliers. High breakdown characteristics are measured by performance on data with a high percentage (10-25%) of outliers. High efficiency is measured by comparing a robust technique's performance against least squares using data containing no outliers. Lastly, the property of bounded influence is typically measured by performance on data containing high leverage point outliers. In total, five datasets are used for this study. Three datasets are used to test the desirable properties and two additional datasets containing multiple point clouds are added to determine the impact of poor leverage measures on GM-estimates. Table 4.4 describes the composition and purpose of each dataset.

Table 4.4. Dataset Description

Dataset	Number of Variables	Sample Size	Number of Leverage Points	Leverage Point Distribution and Location	Outlier Location **X**-space	Outlier Number and Percentage	Property Tested
1	6	40	0	Single - Axis	N/A	0 / 0%	Efficiency
2	6	40	8	Single - Axis	Interior	10 / 25%	Breakdown
3	6	40	8	Single - Axis	Exterior	8 / 20%	Bounded Influence
4	2	16	4	Cloud - Diagonal	Exterior	3 / 20%	Bounded Influence and Leverage
5	2	16	4	Cloud - Off-Line	Exterior	3 / 20%	Bounded Influence and Leverage

Decisions regarding the location of the high leverage points and the magnitude of outlier errors are made based on the number of regressors, sample size, and results from previous pilot studies on this issue. The location of the high leverage points and the magnitude of the errors for each dataset are described in Table 4.5. Error direction (+/-) is selected randomly for all points, including outliers. As a result, it is possible for some datasets to contain outliers all in the same direction, which can provide a challenging scenario for least squares and some of the robust estimation methods. The performance measure calculated for each run in the experiment is the mean square error estimation (MSEE). Variation between runs is obtained using random normal variates for the "good" errors, and varying magnitude and sign on the outliers. A technique's overall performance on a type of dataset is determined by the *average* MSEE of a number of runs. Fifty runs are performed and averaged for each dataset type in order to obtain AMSEE values with acceptably small variances.

Table 4.5. High Leverage Point Location and Error Magnitudes

Dataset	Number of Leverage Points	Leverage Point Location	Leverage Point Distance from Design Center	Outlier Location (X-space)	Outlier Number and Percentage	Outlier Error Magnitude
1	0	N/A	N/A	N/A	0 / 0%	N/A
2	8	Axes (8 of 12 possible)	8, 10, 12, 14 (2 of each)	Interior	10 / 25%	6, 7, 8, 9, 10 (2 of each)
3	8	Axes (8 of 12 possible)	8, 10, 12, 14 (2 of each)	Exterior	8 / 20%	7, 8, 9, 10 (2 of each)
4	4	Cloud - Diagonal	(6, -6)	Exterior	3 / 20%	6, 7, 8
5	4	Cloud - Off-Line	(7, -4)	Exterior	3 / 20%	6, 7, 8

4.5.3 Location of the High Leverage Clouds

The purpose of testing technique performance against Datasets 4 and 5 (DS4 and DS5) is to determine the impact of misspecified leverage distances on final GM-estimation. As noted previously, certain multiple point cloud locations can cause difficulties for each of the proposed measures of leverage. The dataset containing the multiple point cloud located in-line with a diagonal of the cube points (DS4) reveals the inability of all three methods to identify the cloud as high leverage. The methods differ in terms of the degree of misidentification. The MVE distances both mask (makes outliers appear to be inliers) and swamp (makes inliers appear to be outliers), while the hat diagonals and KW weights tend to only mask. In DS5, the MVE distances correctly identify the cloud as high leverage points, while the other two methods continue to mask (see Table 4.6).

Table 4.6. Measures of Leverage for Datasets 4 and 5

	DS4 - Diagonal Cloud		DS5 - Off-Line Cloud	
Leverage Estimation Method	Cube (Inlier)	Cloud (Outlier)	Cube (Inlier)	Cloud (Outlier)
MVE Squared Distances	0.8 - 102.6	0.8 - 11.3	2.0 - 8.0	49.2 - 52.9
Hat Diagonal Values	0.06 - 0.25	0.22 - 0.26	0.07 - 0.24	0.23 - 0.24
KW Weights	5.6 - 15.2	17.4 - 18.6	6.1 - 12.6	18.7 - 19.4

Based on this information, it is expected that the techniques using MVE distances will fit poorly on DS4 and well on DS5. Techniques employing hat diagonals and KW weights should perform reasonably well on both DS4 and DS5.

4.5.4 Conducting the Experiment

The code used to generate the experiments is written and executed in S-PLUS. This screening study requires 2750 (11 x 5 x 50) robust technique estimations, each consisting of initial estimates, followed by estimates of leverage, and most often requiring an iterative convergence scheme to obtain the final coefficient values. To increase the efficiency of the program, redundant estimation is avoided. For example, several of the techniques use Least Trimmed Sums of Squares (LTS) as the initial estimate. Understanding that this estimator is computationally rigorous, it will only be computed once and the LTS coefficient estimates will be directly input into the required techniques. The same approach will be used for leverage distance measures, which will also save tremendous amounts of processing time. The approach also eliminates the variability between techniques due to estimate approximations. For several of the initial estimation techniques, the number of computations required for an exhaustive search and exact solution is not feasible for even moderate size problems (6 regressors and 40 observations require 18 million subset evaluations). This condition necessitates a random subsample approach be used, resulting in approximations for *S*-estimates, LMS, LTS and MVE leverage estimates.

4.5.5 Experiment Analysis

Analysis of the results consists of a detailed study of the AMSEE values for each robust technique and dataset combination. Unusually large or small AMSEE values were immediate candidates for investigation, first to determine whether the technique is operating properly, and second to determine the reason for the strength or weakness of the estimation method.

Particular attention is paid to technique performance on DS2 and DS3. DS2 challenges the techniques' ability to handle several interior X-space outliers and DS3 locates the outliers in the high leverage points to determine the techniques' ability to bound the influence. Of the five DS scenarios, least squares performs the worst on DS2 and DS3. GM-estimator is studied relative to DS2 and DS3 in terms of its ability to obtain good initial estimates and to properly downweight the appropriate outlying observations. Proper downweighting is a function of the choice of leverage measures, the computation of the π-weights, the choice of ψ-function (and tuning constant), and convergence technique used. Table 4.7 shows the AMSEE values for each technique / dataset combination in the experiment. The notes provided for the table discuss technique characteristics contributing to the strong or weak performance of a technique on a particular dataset. A more complete discussion of each technique's overall performance is provided following the notes.

Table 4.7. Bounded Influence Technique Average Mean Square Error of Estimation (AMSEE)

	AMSEE											
Dataset	**LS**	**GMWA**	**GMCH**	**GMMZ**	**GMWA1**	**GMCH1**	**GMCH2**	**GMCH3**	**GMMZ1**	**GMMZ2**	**GMNP1**	**GMNP2**
Normal Error	0.16	0.16	0.24[1]	0.25[2]	0.16	0.19	0.17	0.17	0.16	0.18	0.18	0.17
Int. X-space	0.78	1.62[3]	0.20	0.28	1.52[3]	0.26	0.35	0.35	0.94[4]	0.28	0.44	0.58
Ext. X-space	2.19	1.26[5]	0.82	0.45[7]	0.62	0.89	1.06	0.55	1.09[6]	1.16	0.54	0.49
Diag- Cloud	0.31[8]	0.31	0.43	0.39	0.30	1.57[9]	1.48[9]	65.92[10]	0.30	1.07[9]	0.30	0.29
Off-Line Cloud	0.32[8]	0.37	0.36	0.41[2]	0.34	0.25[11]	0.25[11]	0.25[11]	0.37	0.26[11]	0.38	0.36[12]

Performance Notes

1. High Breakdown Point / Low Efficiency initial estimator is combined with only one Newton GM-iteration resulting in a low efficiency final estimate
2. Significant downweighting of many of the non-outlying observations

3. Uses the Huber ψ-function which is an improvement over Tukey Biweight ψ-function, but still downweights the high leverage points too much
4. Good initial estimate, but converges to a poor final estimate. The final w_i show insufficient interior **X**-space outlier downweighting and too much high leverage point non-outlier downweighting
5. Poor initial estimate (least squares)
6. Tuning constant is set at too large a value, resulting in an estimator that is not sensitive enough to outliers. The tuning constant is modified appropriately in GMNP1 causing a 50% reduction in AMSEE
7. Good results, but too many "good" points are being downweighted
8. Least squares is not adversely impacted by these datasets (DS4 and DS5) because most run variations consist of a combination of positive and negative outliers, so the LS fit is to the middle, exactly where the true observations are located
9. Poor MVE estimates of leverage
10. Poor MVE estimates of leverage combined with a hard redescending ψ-function results in severe downweighting of many "good" observations and moderate downweighting of outliers
11. Accurate estimates of leverage using the MVE approach, results in the best parameter estimates for this dataset
12. Krasker-Welsch weights do not properly estimate leverage, but the final estimates are not unreasonable

4.5.6 Technique Description and Performance Comments

GMWA

Description - Walker's original proposal consists of a least squares initial estimate. The GM portion involves a Schweppe type objective function using a DFFITS-type argument to the ψ-function. Possible enhancements to this technique, suggested by Walker, are incorporated in the alternative technique GMWA1.

Strengths - This technique performed well on DS1, indicating high efficiency. Use of the hat diagonals as indicators of leverage resulted in acceptable performance on the cloud outlier datasets.

Weaknesses - This estimator performed poorly on tests for high breakdown (DS2) and bounded influence (DS3). It finished last on DS2 (25% interior **X**-space outliers), well behind least squares (see note 3 above). Insufficient downweighting of the exterior **X**-space outliers in DS3 resulted in only moderate improvement over least squares. A complete experiment was also performed on a modified GMWA using the hard redescending Tukey's biweight ψ-function in place of the monotone Huber ψ-function. The AMSEE results are (0.18, 3.03, 0.96, 0.42, 0.41) for DS1 through DS5. Comparing these numbers with the GMWA column for DS1-5 (0.16, 1.62, 1.26, 0.31, 0.37) indicates that the Huber ψ-function performs better in all but DS3, the exterior **X**-space outliers dataset. The improved performance using the Huber ψ-function confirms Walker's suggestion to use a redescending ψ-function only in combination with a good initial estimate.

GMCH

Description - Coakley and Hettmansperger's proposal that claims to have high efficiency, high breakdown and bounded influence.

Strengths - This technique is a good overall performer. It is the clear winner in terms of breakdown, having the lowest AMSEE on DS2. It does improve considerably on least squares on the six-variable exterior **X**-space outlier dataset (DS3). Although this technique employs the MVE estimator distances as leverage measures, disastrous results are avoided in DS4 by using only one iteration of Newton's method.

Weaknesses - The high efficiency claim is not well substantiated on DS1, where this technique had an AMSEE 51% larger than least squares. This result is not surprising knowing that the initial estimate, LTS, has low efficiency and only one Newton step is used to move towards a more efficient result. The one-step solution also prevents this technique from outperforming the least squares fit in DS5, the off-line multiple point cloud dataset.

GMMZ

Description - Marazzi's proposal is a method that combines the suggestions of Krasker and Welsch (1982) and Hampel et al. (1986). The initial estimate and measures of leverage do not necessarily have high breakdown. The measures of leverage are used in both the most B-robust initial estimates, as weights in the weighted LAV estimator, and in final estimation, as weights in the IRLS convergence routine.

Strengths - Using the performance metric indicating the percentage over minimum AMSEE (Table 4.9), this technique is the best. It performs surprisingly well in high breakdown situations (ranked third in DS2), and is the best in dealing with bounded influence, ranked first in DS3.

Weaknesses - Although this technique tends to perform well in the presence of outliers, it performs the worst of the robust techniques under normal error conditions. Analysis of the final weights indicates that severe downweighting of non-outlier observations results in low efficiency relative to least squares (Table 4.8). In addition, the superb performance on DS2 and DS3 is tempered by the fact that the technique causes large downweighting of many of the observations, including "good" points.

Table 4.8. Example of Final Weights for GMMZ for the Normal Error Dataset (DS1)

Case	Weight	Case	Weight	Case	Weight	Case	Weight
1	0.08	11	0.39	21	1.00	31	0.12
2	0.20	12	0.36	22	0.27	32	0.75
3	0.49	13	0.16	23	0.07	33	1.00
4	0.13	14	0.18	24	0.09	34	0.17
5	1.00	15	1.00	25	0.06	35	1.00
6	0.51	16	0.80	26	0.18	36	0.05
7	0.26	17	0.13	27	0.26	37	1.00
8	0.72	18	0.22	28	0.37	38	0.50
9	1.00	19	0.27	29	0.11	39	0.92
10	0.25	20	0.21	30	0.16	40	0.07

Both Krasker and Welsch (1982) and Walker (1984) consider the amount of overall downweighting a significant factor in the effectiveness of a GM technique. The preferred estimator is the one that downweights only the discrepant observations. Any unnecessary downweighting is not only undesirable, but results in lower efficiency relative to least squares under normal errors. Krasker and Welsch decided to compare their GM proposal (very similar to GMMZ) to *M*-estimation only after determining the bound on sensitivity (measured by a tuning constant) such that the average final weights were equal. The resulting tuning constant was significantly larger

$(1.71 \cdot \sqrt{p})$ than the constant proposed by Marazzi $(1.05 \cdot \sqrt{p})$. Krasker and Welsch calculate the efficiencies of several tuning constant values using a simple location model. The efficiency of the Marazzi tuning constant is less than 80%.

Walker (1984) compared robust estimates using an inefficiency statistic that measures the percentage of downweighting of the final observations. This measure is computed as

$$INEFF = \left(1 - \frac{\sum w_i}{n}\right) \times 100$$

where w_i are the final weights and n is the number of observations. When comparing results of GM-estimates, smaller inefficiency values are desired. This statistic can be used to show that the GMMZ method has a high inefficiency ratio relative to other robust techniques.

GMWA1

Description - This alternative to the Walker proposal implements his suggestion to use a robust initial estimate in place of least squares. The technique used here for an initial estimate has both high breakdown and moderate efficiency. *S*-estimation provides both the initial estimate and the estimate of scale. Because the Tukey ψ-function performed poorly, a Huber ψ-function is used.

Strengths - The high efficiency from GMWA is maintained along with improved performance in each of the other four scenarios. Most significant improvement is noted in DS3, the high leverage outlier scenario. With the exception of the high breakdown condition, this estimator performs well.

Weaknesses - The primary weakness of this estimator is its inability to provide robust estimates in the presence of many outliers. Dataset 2 contains 25% outliers and GMWA1 has an AMSEE nearly double that of least squares, which is unacceptable. Walker also suggested the possibility of using a redescending ψ-function if a good initial estimate is obtained. Implementing this suggestion by replacing the Huber ψ-function with Tukey's biweight resulted in performance similar to using Walker's original proposal with a redescending ψ-function. The AMSEE results are (0.18, 2.68, 0.47, 0.39, and 0.35) for DS1 through DS5. Encouraging results are again observed regarding efficiency (DS1) and estimation in the presence of high leverage outliers (DS3). Unfortunately, the results for the high breakdown, interior **X**-space outliers is again disastrous. The AMSEE of 2.68 is over four times larger than the AMSEE for least squares (0.65) in that experimental run.

GMCH1

Description - The same components of GMCH are used, but the convergence technique is modified to incorporate fully iterated IRLS. The intent of this modification is to hopefully increase the efficiency while not decreasing the breakdown performance substantially.

Strengths - The intent of the modification was fulfilled regarding increased efficiency. The AMSEE for DS1 decreased from 0.24 to 0.19. Breakdown did not decrease substantially as the AMSEE for DS2 increased only from 0.20 to 0.26. Full iteration of the GM objective is helpful if the corresponding π-weights are accurate, as in DS5. In this case, the AMSEE decreased from 0.36 to 0.25.

Weaknesses - Fully iterating the GM objective can be damaging if the MVE distances used as π-weights incorrectly identify high leverage points as inliers and inliers as high leverage points. The AMSEE for DS4 grew from 0.43 in GMCH to 1.57 in GMCH1. Interestingly, performance in terms of bounded influence (DS3) is better using a single-step GM convergence (0.82) than the fully iterated IRLS (0.89).

GMCH2

Description - GMCH1 is modified by replacing the LTS initial estimate with an *S*-estimator. *S*-estimates have the same breakdown (50%) and slightly higher efficiency than LTS. The initial *S*-estimate of scale is also used throughout the technique. The intent of this modification is to improve the efficiency of the final estimate while not decreasing breakdown.

Strengths - Further improvement on final estimate efficiency is made primarily because the initial estimate is more efficient. Performance remains strong in DS5.

Weaknesses - Performance decreases in terms of breakdown and bounded influence (DS2 and DS3) relative to GMCH1. This technique fails to improve on the poor performance against the diagonal cloud dataset (DS4).

GMCH3

Description - GMCH2 is modified in terms of the ψ-function. The Tukey biweight function replaces the Huber ψ-function.

Strengths - This technique performs substantially better than any of the GMCH alternatives on the high leverage outlier dataset because of the more severe downweighting property of the ψ-function.

Weaknesses - As described in performance summary note 10 above, poor MVE estimates combined with a hard redescending ψ-function can produce meaningless coefficient estimates.

GMMZ1

Description - Due to the heavy downweighting behavior of GMMZ, an attempt was made to scale the KW weights so that nonoutliers do not receive leverage downweighting. A simple scaling approach consists of multiplying each π-weight (equal to the inverse of the KW weights) by the median KW weight in the dataset. This scaling factor results in a π-weight of 1 for the median KW weight observation and a fraction less than one for all other observations.

Strengths - The modified π-weights successfully reduced the large downweighting of nonoutliers, substantially increased the LS efficiency and caused improved performance in DS1 and DS4.

Weaknesses - Significant reductions in performance on DS2 and DS3 indicates that perhaps the tuning constant requires modification. This change is adopted in GMNP1.

GMMZ2

Description - The KW weights used as measures of leverage are replaced by the MVE distances along with the suggestion for MVE π-weights. The MVE weights have a high breakdown property not present in KW weights. Expect improvements over GMMZ1 in terms of breakdown performance. The tuning constant is changed to reflect the altered leverage measure. This technique is identical to GMCH2 except for the initial estimate.

Strengths - Introduction of MVE distances into this technique results in anticipated improved performance over GMMZ1 on the high breakdown dataset (DS2) and the off-line cloud dataset (DS5).

Weaknesses - Like GMMZ1 and GMCH2, this technique does not perform as well as some of the others on the high leverage outlier dataset (DS3).

GMNP1

Description - This technique is identical to GMMZ1 except for the tuning constant. Because the π-weights are scaled, it makes sense to remove the number of parameters term (p) from the tuning constant and use the Huber 95% efficiency value, similar to the MVE tuning constant approach.

Strengths - This approach improves substantially on the weaknesses of GMMZ1 (on DS2 and DS3) without changing the good performances on the other datasets.

Weaknesses - Slight decreases in efficiency are observed (DS1).

GMNP2

Description - GMNP1 is modified by substituting *S*-estimation for the initial estimate in place of the most B-robust estimate.

Strengths - This technique has no significant weaknesses. Efficiency improves over GMNP1 slightly as well as the performance on DS3, DS4 and DS5.

Weaknesses - Performance on the high breakdown interior X-space outlier dataset could be improved.

4.6 Performance Comparison Summary

Determining the best overall performing GM techniques is accomplished by evaluating several indicators of AMSEE performance. One indicator of AMSEE performance is the relative AMSEE rank of each technique on each dataset. The AMSEE values are ranked lowest to highest by dataset, and the ranks are summed for each technique. A lower ranking indicates better AMSEE performance. The technique with the smallest sum of ranks is then considered, by this indicator, the best overall performer.

Another indicator of performance is the standard deviation of the ranks, which indicates the stability of the technique performance. Smaller rank standard deviations are preferred. This indicator is most valuable when comparing techniques with similar summed ranks.

A third indicator of AMSEE performance involves accounting for the differing AMSEE ranges among techniques within a particular dataset. For instance, the spread between the lowest and highest AMSEE for DS1 is 0.09, while the spread for DS2 is 1.42. Being ranked last in DS1 may not be as harmful as being ranked last in DS2. One method for capturing this spread within datasets is to compute, for each technique/dataset combination, the percent above the minimum AMSEE. For example if the smallest AMSEE is 0.20 for DS2, the percent above minimum AMSEE for a technique with AMSEE of 0.40 is 100%. Summing these percentages is another indicator of AMSEE performance. The results of these techniques in terms of their performance is displayed in Table 4.9.

Table 4.9. Indicators of AMSEE Performance for GM-Estimation Techniques

Performance Rank											
Dataset	**GMWA**	**GMCH**	**GMMZ**	**GMWA1**	**GMCH1**	**GMCH2**	**GMCH3**	**GMMZ1**	**GMMZ2**	**GMNP1**	**GMNP2**
Normal Error	1	10	11	3	9	4	5	2	7	8	6
Int. X--space	11	1	3	10	2	6	5	9	4	7	8
Ext. X-space	11	6	1	5	7	8	4	9	10	3	2
Diag- Cloud	5	7	6	2	10	9	11	3	8	4	1
Off-Line Cloud	8	6	11	5	2	3	1	9	4	10	7
Sum of Ranks	36	30	32	25	30	30	26	32	33	32	24
Overall Rank	**11**	**4**	**7**	**2**	**4**	**4**	**3**	**7**	**10**	**7**	**1**
Rank Std Dev	**4.27**	**3.24**	**4.56**	**3.08**	**3.81**	**2.55**	**3.63**	**3.58**	**2.61**	**2.88**	**3.11**

Percentage Above Minimum AMSEE											
Dataset	**GMWA**	**GMCH**	**GMMZ**	**GMWA1**	**GMCH1**	**GMCH2**	**GMCH3**	**GMMZ1**	**GMMZ2**	**GMNP1**	**GMNP2**
Normal Error	0%	51%	60%	4%	18%	8%	8%	0%	13%	14%	9%
Int. X-space	708%	0%	39%	662%	32%	75%	75%	371%	41%	122%	189%
Ext. X-space	180%	82%	0%	37%	97%	136%	22%	142%	158%	20%	9%
Diag- Cloud	9%	49%	37%	3%	446%	416%	22813%	3%	273%	4%	0%
Off-Line Cloud	45%	41%	64%	35%	0%	1%	0%	48%	2%	52%	43%
Sum of Percents	943%	223%	199%	741%	593%	636%	22918%	565%	487%	213%	251%
Overall Rank	**10**	**3**	**1**	**9**	**7**	**8**	**11**	**6**	**5**	**2**	**4**
Sum of Overall Ranks	**21**	**7**	**8**	**11**	**11**	**12**	**14**	**13**	**15**	**9**	**5**

4.7 Conclusions

It is important to note that the purpose of this study is not to determine the best single GM technique. More datasets would be required to fully test each technique and a careful weighting system is required to determine the importance of each dataset in the overall performance measure. The purpose in this study is to identify a few promising techniques to be used in a more comprehensive study. Also, although some drawbacks have been noted in the characteristics of some of the techniques, such as too much overall downweighting, no technique will be eliminated only for these reasons.

The best performing techniques based on the indicators in Table 4.9, are GMCH, GMMZ, GMNP1, and GMNP2. The first two techniques are the original proposals of Coakley and Hettmansberger (GMCH) and Marazzi (GMMZ). The two most promising alternatives are the more modified alternatives GMNP1 and GMNP2. Both alternatives use Krasker-Welsch weights for leverage and the modified, scaled π-weights. The two alternatives differ in their choice of initial estimates and estimates of scale. GMNP1 uses the most B-robust initial estimator with the MAD for scale, and GMNP2 uses *S*-estimators for both the initial and scale estimates. The weaknesses of the original proposals appear to be in efficiency, while the alternatives may either have trouble with interior **X**-space outliers or with breakdown. Continued testing of these methods in the following chapters will further define their strengths and weaknesses.

Chapter 5

A Capability Assessment of Robust Regression Methods

5.1 Introduction

Robust estimation methods in regression are designed to enable the model builder to fit equations to the majority of the data in the presence of outliers. Outliers can arise for many different reasons and can appear in many different forms. Reasons for outliers include coding or computational errors, copying errors such as misplaced decimals, observations that are not necessarily part of the population being studied, machine or equipment failure, or even transient effects. Outliers can occur as single observations, in groups, or scattered throughout the data. Outliers can be located in the interior or exterior regressor space region. Regardless of their configuration, it is the goal of robust techniques to locate these outliers, appropriately downweight their influence and fit an equation to the remainder of the observations. Because many regression estimation routines are in place in industry which automatically import data and fit equations without performing diagnostics such as outlier detection, it is equally important that robust methods perform well on data with and without outliers.

This seemingly reasonable request is often an insurmountable challenge for many robust methods, especially if the goal is to accurately fit to the majority of the data for a *variety* of outlier conditions. Many of the proposed techniques can handle certain outlier configurations very well, but provide poor estimates for other outlier scenarios. For this reason, no robust method has been developed that clearly estimates well under all possible outlier and nonoutlier conditions.

Two properties of robust estimators are of interest when outliers are present in the data. The first property, **breakdown**, deals with the number of outliers present in the data and their affect on technique estimation. Least squares estimation can be rendered useless by the presence of a single outlier. Breakdown is a measure of the percent of outliers in the data before an estimation technique is no longer reliable. By this definition least squares has a zero percent breakdown. Obviously, high breakdown point estimators are desirable. Some robust techniques have breakdown points as high as 50%.

The second desirable property, **bounded influence**, concerns the location of the outlier(s). Least squares estimation is affected more by observations that are in the exterior of the regressor or **X**-space region. These observations have considerably more influence on the least squares equation. The objective of robust techniques in this regard is to bound the influence of these exterior points. Some robust methods can effectively reduce the impact of interior point outliers but cannot reduce or bound the influence of exterior point outliers.

If outliers are *not* present in the data, robust techniques should be nearly as **efficient** as least squares, which is the best linear unbiased estimator. In fact, robust methods are often compared with least squares under normally distributed error situations (with no outliers present). Relative efficiency is a measure of robust technique's estimation ability relative to least squares for normal error data. An asymptotic (large sample) relative efficiency ratio can be calculated for most robust techniques. The properties of breakdown and efficiency are often competing in the sense that it is difficult to design robust methods that have both high breakdown and high efficiency. In the following section, robust techniques are grouped and compared relative to the properties of breakdown, bounded influence and efficiency.

The purpose of this chapter is to develop and conduct a comprehensive evaluation of competing robust regression methods. These competing methods consist of a group of top performing existing techniques and a group of the best performing proposed methods from this body of work (Chapter 4). The techniques will be evaluated in each of two experiments designed to fully test their estimation capabilities. Datasets will be developed to test methods under nonoutlier condition and various combinations of outlier percentage, magnitude and location. Performance measures used to compare techniques include the mean square error of estimation and the mean square inefficiency ratio, which indicates relative improvements over least squares. Techniques are also compared in terms of their relative performance ranks across outlier configurations. The results of this study are intended to provide the user with the best overall performing robust method.

5.2 Selected Regression Techniques

The performance of several respected robust regression techniques are measured and compared against each other and with least squares estimation in a study designed to determine the method with the best overall performance. The robust techniques consist of a series of established techniques, recently published and promising techniques, and some new GM-estimation proposals recently developed and tested. By adding least squares estimation to the analysis, it not only reveals the relative "pull" that outliers may have on estimation, but it also provides a benchmark for evaluating each robust alternative's relative improvement. The best overall performing technique, determined by the use of a number of performance statistics, is a method that consistently outperforms least squares and ranks among the top robust estimators in terms of overall error of estimation relative to the true model coefficients. Additional consideration is given to the consistency of the performance as measured by the standard deviation of the relative ranks. Top performing robust techniques with smaller rank standard deviations indicate repeated high performance with no or few vulnerabilities. Large rank standard deviations for top performers are indications of techniques that may not estimate well under certain conditions.

The techniques considered for this study will be briefly discussed. Previous experiments have been performed to study a more exhaustive list of robust techniques and have resulted in a reduced set of methods. The robust methods used in this study consist of top performing techniques representing different classes of robust estimators. The classes of estimators include: a) high efficiency, b) high breakdown, and c) multiple property techniques. Table 5.1 outlines the robust techniques used in the initial experiment, the original reference and the associated desirable properties.

The table shows that the robust techniques consist of two high efficiency methods, two high breakdown methods and six multiple property techniques that combine two or three characteristics of efficiency, breakdown and bounded influence. The descriptive characteristics are not guarantees of high performance under certain conditions, they are strictly theoretical properties associated with the estimator. In fact, there is some debate concerning the expected empirical performance for some estimators that theoretically are high efficiency and high breakdown. Thus, one of the primary purposes of these experiments is to compare the theoretical properties of techniques with their actual performance on simulated data designed to test the properties.

Table 5.1. Robust Regression Techniques and Associated Properties - Experiment 1

Technique	Origin	Properties
Single Stage		
M-estimation	Huber (1973)	Efficiency
Most B-Robust	Hampel et al. (1986)	Efficiency
Least Trimmed Sums of Squares (LTS)	Rousseeuw (1983, 1984)	Breakdown
S-estimation	Rousseeuw and Yohai (1984)	Breakdown
Multiple Stage		
MM-estimation	Yohai (1987)	Efficiency, Breakdown
GM-estimation *	Coakley and Hettmansperger (1993)	Efficiency, Breakdown, Bounded Influence
GM-estimation *	Marazzi (1993)	Efficiency, Bounded Influence
GM-estimation *	new proposal	Efficiency, Bounded Influence
GM-estimation *	new proposal	Efficiency, Bounded Influence
GM-estimation *	new proposal	Efficiency, Breakdown, Bounded Influence

* GM-estimation was originally proposed by Mallows (1975), Hill (1977), Hampel (1978) and Krasker (1980). The origins mentioned for GM-estimates above are the authors of specific variations.

5.2.1 Technique Selection Rationale

The robust estimation methods contain well-known established techniques as well as recently developed, promising alternatives. *M*-estimation is included due to its popularity and known success in modeling data with interior point influence. The most-B robust method, which is actually a least absolute value technique weighted by measures of leverage, is regarded as a promising technique by Hampel et. al. (1986) and Marazzi (1993). Most B-robust is also used as an initial estimate in one of the candidate GM-estimation techniques. Least Trimmed Sums of Squares (LTS) is included because of its popularity as a high breakdown point technique that has slightly higher efficiency than its high breakdown counterpart, Least Median of Squares (LMS).

LTS is also included for its use as a GM technique initial estimate. *S*-estimation is a high breakdown method with higher efficiency than LTS, and is used as an initial estimate in two candidate multi-stage estimators: one of the proposed GM techniques and in MM-estimation (Yohai 1987). MM-estimation is a high performance two-stage technique that has both high efficiency and high breakdown.

Four GM-estimation techniques were originally selected from the experiment conducted in Chapter 4 to identify the best performing GM-estimation techniques. Two of these techniques are proposals from the literature. One of the proposals originated from Coakley and Hettmansperger, who developed a technique with high efficiency, high breakdown and bounded influence. The other published proposal is that of Marazzi, who proposes a method using the most B-robust initial estimator, and KW weights with either Newton or IRLS fully iterated convergence. The other two GM-estimation proposals are alternatives suggested in the previous study. GMNP1 is a method using the most B-robust initial estimate followed by GM fully IRLS convergence. The π-weights and ψ-function cutoff value for this technique are modified from Marazzi's proposal. Finally, the GMNP2 method uses *S*-estimation as the initial step and the same final stage as GMNP1 to obtain final coefficient estimates.

As a result of initial runs of this robust method screening design, it was observed that a modification of one of the GM alternatives may outperform the other GM methods. Thus, we developed a fifth GM-estimation technique, GMNP3. This technique requires only a slight modification to GMNP2, which is an *S*-estimate initial step, followed by a fully iterated (IRLS) Schweppe-type objective using KW weights, and Huber ψ-function. After realizing that for some of the datasets the final estimate of GMNP2 was actually converging to a worse than initial solution, it made sense to try implementing a reduced step convergence technique. Several alternative step sizes were tested, including one, two and three-step IRLS convergences. Further analysis revealed that the one-step IRLS approach performed the best. Using the one-step IRLS, the final estimate is really just a weighted least squares using the weights as defined in GMNP2. Initial tests show that this method outperforms its fully iterated counterpart because it tends to take big steps in the correct direction (away from the initial estimate toward an improved solution) and small steps in the incorrect direction, toward a worse solution. In certain instances it is also observed that the fully iterated method takes the first IRLS step toward an improved solution and converges to an estimate worse than the first IRLS step. A brief discussion of the mechanics of the robust methods is provided in the following section. A more detailed description of these methods is provided in Chapter 2.

5.2.2 *M*-estimators

The class of estimators called *M*-estimators was first introduced by Huber (1973) and different variations of this approach by other authors quickly followed. *M*-estimators are probably the most widely known and applied robust estimation techniques. The *M*-estimators are based on the idea of replacing the sum squared residuals by a more gradually increasing function of the residuals $\rho(r)$, where ρ is a symmetric function with a unique minimum at zero.

$$\min_{\beta}\sum_{i=1}^{n}\rho\left(\frac{e_i}{s}\right)=\min_{\beta}\sum_{i=1}^{n}\rho\left(\frac{y_i-\mathbf{x}_i'\hat{\beta}}{s}\right) \tag{5.1}$$

M-estimators are maximum likelihood estimators in which the function ρ is related to the likelihood function for an appropriate choice of the error distribution.

Taking the first partial derivatives of (5.1) with respect to β and setting the result equal to $\mathbf{0}$, as

$$\min_{\beta}\sum_{i=1}^{n}\psi\left(\frac{e_i}{s}\right)=\min_{\beta}\sum_{i=1}^{n}\psi\left(\frac{y_i-\mathbf{x}_i'\hat{\beta}}{s}\right)\mathbf{x}_i=\mathbf{0} \tag{5.2}$$

where $\psi(u)=\frac{\partial}{\partial u}\rho(u)$, resulting in the necessary condition normal equations. With *M*-estimation objective functions, $\psi(u)$ is not linear so that (5.2) defines a nonlinear system of equations which requires an appropriate nonlinear iterative estimation technique.

The $\psi(u)$ function controls the weight given to each residual and is very important in determining the robust and efficiency properties of the estimator. Although a number of popular ψ-functions have been developed, the proposals are of two types: monotonic and redescending. The Huber function (Huber 1964), is an example of a monotone ψ-function defined as

$$\psi(u)=min(c_H, max(u,-c_H)) \tag{5.3}$$

which results in down-weighting the large residuals compared to least squares. Other ψ-functions redescend with increasing residual magnitude. The bisquare or biweight function of Beaton and Tukey (1974), is defined as

$$(u)=\begin{cases} u(1-(u/c_B)^2)^2 & for|u|\le c \\ 0 & for|u|>c \end{cases} \tag{5.4}$$

The c terms in both equations refer to tuning constants chosen to achieve desired efficiencies. The values c_H=1.345 and c_B=4.685 for the Huber and biweight ψ-functions respectively achieve 95% efficiency compared to the least squares estimator in the model when the errors are actually normally distributed.

5.2.3 Most B-Robust Estimators

The most B-robust estimator is of the class of *M*-estimate models. The objective of this estimator is to find the minimum of the weighted absolute value of the residuals

$$\min_{\beta} \sum_{i=1}^{n} w_i \left| y_i - x_i \hat{\beta} \right| \tag{5.5}$$

where $w_i = 1/|z_i|$ and the z_i are measures of leverage which can be computed using *M*-estimates of covariance. The z_i can be interpreted as the distance a point lies from the center of mass in the **X**-space. The most B-robust estimator is actually just a weighted LAV estimator. This approach is not considered high breakdown and does not have bounded influence.

5.2.4 Least Trimmed Sum of Squares

The least trimmed squares (LTS) approach was developed by Rousseeuw (1983, 1984) as a high efficiency alternative to least median of squares (LMS). The LTS estimator is given by

$$\min_{\beta} \sum_{i=1}^{h} (e^2)_{i:n} \tag{5.6}$$

where $(e^2)_{1:n} \leq (e^2)_{2:n} \leq \ldots \leq (e^2)_{n:n}$ are the ordered squared residuals and h is the number of residuals included in the calculation. This approach is similar to least squares except the largest α squared residuals are not used (trimmed sum) in the summation, allowing the fit to avoid the outliers. This approach converges at a rate similar to the *M*-estimators. It is also equivariant and the breakdown point is 50% when h=n/2. According to Rousseeuw and Leroy, the main disadvantage of LTS is the large number of operations required to sort the squared residuals in the objective function. Another challenge is deciding the best approach for determining the initial estimate.

5.2.5 *S*-estimators

This technique consists of a class of estimates based on the minimization of a robust *M*-estimate of the residual scale. They are defined by minimization of the dispersion of the residuals:

$$\min_{\beta} \ s\left(e_1(\beta),\cdots,e_n(\beta)\right) \tag{5.7}$$

The dispersion function $s\left(e_1(\beta),\cdots,e_n(\beta)\right)$ is found as the solution to

$$\frac{1}{n-p}\sum_{i=1}^{n}\rho\left(\frac{y_i-\mathbf{x}_i'\hat{\beta}}{s}\right)=K \tag{5.8}$$

The constant K may be defined as $E_{\Phi}[\rho]$, where Φ stands for the standard normal distribution. The usual choice for the ρ function is Tukey's biweight function whose derivative is the ψ-function given by (5.4)

The term *S*-estimator is used to describe this class of robust estimation because the scale statistic s is implicitly derived as an *M*-estimate of scale. These estimators have the characteristics of a high breakdown point (up to 50%), and are asymptotically normal. The corresponding asymptotic (relative) efficiency for the normal error model can be calculated for various breakdown combinations of K and c. Efficiency can be increased at the expense of breakdown. The efficiency for a 50% breakdown *S*-estimator is 28.7%, which is significantly higher than the efficiency of the 50% breakdown LTS estimator, which is 7.1%.

5.2.6 MM-estimators

Yohai (1987) introduced multi-stage estimators called MM-estimates, which combine high breakdown with high asymptotic efficiency. MM-estimates are computed using a three-stage procedure. The first step involves the computation of an initial estimate with high breakdown properties. Yohai suggests using the *S*-estimate for this initial estimator. The second stage is used to compute an *M*-estimate of the errors scale using the initial step *S*-estimate residuals. Lastly, in the third stage an *M*-estimate of the regression parameters based on an appropriate redescending ψ-function is computed.

Since Yohai's (1987) original proposal, refinements have been suggested by several authors including Ruppert (1992) and Yohai et al. (1991). The algorithm detailed below includes these refinements and the suggested implementation is provided by Marazzi (1993). This implementation includes the test for bias suggested by Yohai et al. which uses a student's T test statistic to determine whether the bias in the final estimate may be unacceptably high and perhaps the initial estimate should be used for exploratory purposes.

5.2.7 GM-estimators

A robust technique that attempts to downweight the high influence points as well as large residual points is GM-estimation. The GM-estimators are solutions to the normal equations formed by

$$\sum_{i=1}^{n} \pi_i \psi\left(\frac{y_i - \mathbf{x}_i' \hat{\beta}}{s\pi_i} \right) \mathbf{x}_i = \mathbf{0} \tag{5.9}$$

where, for appropriate values of π_i the GM-estimator, nicknamed the bounded influence estimator, can downweight outliers with high leverage points. The variation of GM-estimator described here was developed by Schweppe (see Hill 1977). The other common type of GM-estimator was proposed by Mallows (1975). The distinction between these two types is that the Mallows estimator does not have the π-weight in the denominator of the ψ-function. Both types of objectives have the effect of downweighting leverage points, but the Schweppe weighting scheme downweights only if the residuals are large.

Iterative techniques such as Newton's method or IRLS are used to solve the system of equations given by (5.9). Several suggestions for the π-weights have been made that include different measures of leverage. Common approaches to measuring leverage include using the hat diagonals, *M*-estimates of covariance, and robust distances using the minimum volume ellipsoid estimator (MVE). GM-estimators posses the same efficiency and asymptotic distributional properties of *M*-estimators. The breakdown point of the GM approach improves on the $1/n$ value of *M*-estimation, but is still not considered a high breakdown point estimator. The breakdown point is a function of the number of variables p, and is no greater than $1/p$. This condition can lead to problems in models with many regressors. Also, both *M*-estimation and GM-estimation can be improved by starting with a good initial estimate.

Computation of GM-estimates require two stages of parameter estimation consisting of an initial estimate that provides a good starting point, followed by convergence to the final GM-estimate. Development of a technique requires that a number of decisions be made regarding choices of initial estimators, estimators of scale and leverage, type of π-weight and ψ-function, and type and amount of convergence necessary. Table 5.2 outlines the necessary decisions, provides a description and offers some considerations.

Table 5.2. GM-Estimation Technique Characteristics

GM-Component	Comments
Bounded Influence Objective	The preferred type is that of Schweppe, who proposed an objective that theoretically downweights high leverage points only if the residual is large
Initial Estimate	The intent is to provide a good starting point. High breakdown estimators are typically used
Estimate of Scale	Several high breakdown choices are available including the MAD, the LMS estimate of scale, and the scale output of the initial estimate (from *S*-estimates). The scale estimate can be updated in final estimate iterations, but convergence is not necessarily assured
Estimate of Leverage	Different methods are available. A tradeoff exists between computational ease and the ability to handle clouds of multiple outliers
π-weights	Several different approaches are available corresponding to the type of leverage measure used. Some approaches require inlier/outlier cutoff values
ψ-function	*M*-estimate residual downweighting functions including Huber's *t*, Tukey's biweight and Ramsay's exponential
Tuning Constant (ψ-function)	Depends on the ψ-function and desired efficiency. Sometimes it is also a function of *n* and *p*
Convergence	Nonlinear convergence algorithms include, for example, Newton's method, and Iteratively Reweighted Least Squares (IRLS). Another consideration is the number of iterations so that the initial estimate breakdown property is preserved

Several GM-estimation methods consisting of various combinations of GM-components have been proposed over the last decade. Two of these proposals will be implemented and tested against several datasets. In addition, variations of the proposals that have performed well in Chapter 4 experiments will be tested in an effort to find the best performing technique. The published techniques are proposals by Coakley and Hettmansperger (1993), and Marazzi (1993).

The GM-component descriptions for each of the proposals, including the alternatives developed in this research, is provided in Table 5.3. A variety of components are used among the five proposals. Preliminary GM-estimate analysis shows that the Marazzi method performs well against a variety of outlier conditions. Unfortunately, this method tends to heavily downweight many of the observations, even those that are *not* outliers. The three alternatives, labeled NP1-NP3, are variations of the Marazzi method. The differences between the Marazzi approach and these proposed alternatives are highlighted in **bold** in the table. The first alternative, NP1, is an attempt to reduce the level of downweighting by scaling the leverage distances (z), thereby lessening their impact in the GM objective ψ-function.

$$\pi_i = \underset{j}{\mathrm{med}}|z_j|/z_i$$

As a result of this scaling, the median leverage distance in the dataset receives a π-weight of one and the remaining observations have weights that are a ratio of the median distance. The scaling of the weights also eliminated the need for the tuning constant to be a function of the number of model parameters.

The second GM-estimation alternative, NP2, replaces the robust, efficient most B-robust estimation with a high breakdown point estimator. Of the three most popular high breakdown estimators (LMS, LTS and *S*-estimation), *S*-estimation is the most efficient. *S*-estimation is used as the initial estimate and for the initial estimate of scale. The remaining GM components are unchanged from NP1.

The third alternative is a slight but important modification to NP2. Studies of the behavior of NP2 indicate that in datasets containing interior **X**-space outliers, a good initial *S*-estimate was often made significantly worse in IRLS convergence. This behavior suggests that perhaps a one-step convergence to maintain the desirable initial estimate behavior may improve overall performance. A one-step weighted least squares method is implemented using the GM-estimation weights and ψ-function described in NP1 and NP2.

Table 5.3. Published and Proposed GM-Estimation Techniques

	Technique				
GM Component	**Coakley and Hettmansperger**	**Marazzi**	**NP1**	**NP2**	**NP3**
GM Objective	Schweppe	Schweppe	Schweppe	Schweppe	Schweppe
Initial Estimate	LTS	Most B-Robust	Most B-Robust	*S*-estimator	*S*-estimator
Scale Estimate	$\hat{\sigma}_{LMS}$	MAD	MAD	$\hat{\sigma}_{S-est}$	$\hat{\sigma}_{S-est}$
Leverage Measure	Robust Distance (based on MVE)	Krasker-Welsch weights (z)	KW weights.	KW weights.	KW weights.
π-weight Function	min(1, b/RD^2)	$1/\lvert z\rvert$	med$\lvert z\rvert$ / $\lvert z\rvert$	med$\lvert z\rvert$ / $\lvert z\rvert$	med$\lvert z\rvert$ / $\lvert z\rvert$
ψ-function	Huber	Huber	Huber	Huber	Huber
Tuning Constant	1.345	$1.05 \cdot p^{1/2}$	1.345	1.345	1.345
Convergence Approach	One-Step Newton	Iterated IRLS	Iterated IRLS	Iterated IRLS	One-Step WLS
Properties					
High Efficiency	Yes	Yes	Yes	Yes	Yes
High Breakdown	Yes	No	No	No	Yes
Bounded Influence	Yes	Yes	Yes	Yes	Yes

5.3 Experimental Design - Experiment 1

Previous experiments (Chapter 4) were designed to screen the large number of GM-estimation techniques and find the best performing subset of GM methods which would then be used in a comprehensive robust technique comparison. A subset of five of the best performing GM techniques have been identified and are now combined with the best performing other robust techniques. The GM-estimator screening experiment involved a diverse, but not necessarily exhaustive collection of dataset types that were used for measuring model estimation performance. The purpose of this chapter is to test and compare the most promising robust techniques using Monte Carlo simulation with a comprehensive group of outlier datasets.

Datasets are developed that challenge the techniques' ability to display characteristics of high efficiency relative to least squares, high breakdown, and bounded influence. In addition, some unique outlier configurations are developed and tested, including the multiple point cloud configuration and large error value datasets. The number of model parameters *p* is varied to determine the impacts of larger problems on model estimation accuracy. An obvious reason for varying the number of parameters is that the breakdown of GM-estimators has been shown by Maronna and Yohai (1991) to be a function of the number of model parameters. Specifically, they show that GM-estimation breakdown is at most $1/p$, so the breakdown decreases as the model dimension increases. Other factors that have been demonstrated to significantly impact model estimation in previous studies include outlier location in the design **X**-space (interior or exterior), the percent of outlying observations, and the presence of high leverage points. These model estimation factors are varied in the development of experiments used to test the robust techniques' performance.

Previous approaches to experimentation involved a sequential screening process that has been successful for various reasons. Sequential experimentation reduces the chances of implementing designs larger than ultimately required and also enables the experimenter to use information from earlier analyses to properly design subsequent experiments.

This study contains two sequential robust technique performance experiments. The first experiment is developed to test the best performing robust regression techniques against 16 dataset configurations that contain varied number of parameters, sample sizes, leverage point quantities and locations, and outlier densities and magnitudes. The purpose of this experiment is to understand the estimation behavior of the robust techniques relative to efficiency, breakdown and bounded influence. The findings of the first experiment lead to an improved secondary design and revised technique compilation. The second experiment is a more rigorous test that contains the most important outlier factors relative to robust technique performance.

The second experiment consists of four distinct sub-designs. One design is a two-factor test for robust technique efficiency relative to least squares. The factors are the number of model parameters and an indicator variable representing the presence of leverage points. The second design is a single factor experiment that examines technique performance against multiple point cloud locations. The third sub-experiment is designed to study technique performance against datasets containing outliers several orders of magnitude removed from the rest of the data. The final sub-design contains three factors to test performance regarding breakdown and bounded influence. The factors consist of the number of model parameters, outlier density, and a term representing outlier location and leverage presence.

5.3.1 Experiment 1 Description

Although several dataset developments and tests have been performed to this point, some questions regarding technique performance with outliers grouped in certain locations and at certain densities remain unanswered. Therefore, before performing a structured designed experiment involving a small number of factors at different levels, a final performance evaluation of the techniques is required to investigate possible weaknesses against certain outlier configurations. The characteristics of each dataset will be discussed regarding its outlier layout and the properties of the techniques we are interested in testing.

The three desirable properties of robust techniques; breakdown, bounded influence and efficiency, have been described in detail in previous sections. These properties are related to challenges associated with outliers in many empirical datasets. To describe the characteristics of the datasets in this and future experiments, new terms will be introduced that are related to the desirable properties, but are more descriptive of the data than they are of the estimation property.

Testing for a technique's ability to handle high breakdown will consist of more than observing the percentage of outliers in the data. As previously mentioned, GM-estimators have breakdown points that are a function of the number of parameters. The remainder of the robust techniques have a fixed breakdown point for a set of subsample selection sizes and tuning parameters used. Certain techniques, such as *M*-estimation, have a breakdown of 0% regardless of the number of parameters used. Other techniques including LMS, LTS and *S*-estimation have as high as a 50% breakdown, depending on values for the subsample selection size and tuning constant. Neither of these types of techniques has a breakdown point sensitive to the number of model parameters. Because GM-estimators are sensitive to the number of parameters, their breakdown characteristics will be used as the guideline. Datasets will be characterized as *high outlier density* (HOD) if the percentage of outliers exceeds $1/p$, where p is the total number of model parameters, including the intercept.

Regarding the property of bounded influence, or the ability to reduce the influence of high leverage points, datasets will be characterized as *high leverage outlier* (HLO) if points are included that are high leverage and have high error values. If the dataset contains high leverage points that all have small errors, the dataset will not be characterized as high leverage outlier.

In determining robust technique efficiency, we are interested in learning how robust techniques perform estimating "clean" data, with normally distributed errors. For this test, the dataset errors will be random variates drawn from a relatively small variance normal distribution. The variance of normal errors is selected such that the signal-to-noise ratio of the model is about 100:1. These types of datasets will correspondingly be referred to as *normal error* (NE) data.

The situation not addressed by the desirable technique properties may be one of the most common occurrences of outliers. When outliers are located in the interior **X**-space points in low quantity, most robust techniques should perform well. Robust techniques may all perform better than least squares in such a situation, but relative to one another, there may be large performance differences. These points of high error value that are in the interior or inlying **X**-space position, are often referred to as interior **X**-space outliers. This term however is not always an accurate descriptor because it implies that the resulting outliers have response values that are outside the range of the other responses. It is possible though to have datasets with both interior **X**-space points and high leverage points, but its outliers are located only in the interior **X**-space region. Because the response values for the non-outlying high leverage points often define the interior **X**-space range, the outlying interior points may have large errors but still not be interior **X**-space outliers. For this reason, we will refer to these types of outliers as conditions of *interior point influence* (IPI). This term will be used along with the three other terms to describe the characteristics of the datasets used in this experiment.

5.3.2 Individual Dataset Description

The intent of the first robust technique experiment is to expose the techniques to each of the four dataset characteristics, while focusing primarily on high leverage outliers and high outlier density. Datasets containing different high leverage point configurations were tested including the multiple point cloud scenario. Also developed are datasets containing as much as 25% outliers, which for even moderate dimension models (e.g. 6 regressors) are considered high outlier density using our definition from GM-estimators. The datasets will be discussed using their dataset (DS) number and can be referenced in Table 5.4.

The first dataset (DS1) is strictly designed to test for efficiency relative to least squares. The moderate sized model using six regressors is employed. The axial points are located at the same distance from the center of the cube as the design corner points, so that there are no high leverage points. The signal-to-noise ratio is the similar to previous designs (96:1), resulting from coefficient values of 4.0 and a normal error variance of 1.0. The second dataset (DS2) consists of interior point influence and low outlier density (10%). No high leverage points are included that may further complicate the estimation. DS3 differs only from DS2 in the magnitude of half of the axial points, converting them to high leverage points. This dataset then contains 10% high leverage points and 10% interior point outliers.

Table 5.4. First Experiment Runs for the Robust Technique Performance Comparison

DS	Vars / Sample	Leverage # / %	Leverage Location	Leverage Distance	Outliers # / %	Outlier Location X-Space	Outlier Magnitude	Dataset Characteristics
1	6 / 40	0 / 0	N/A	N/A	0 / 0	N/A	N/A	Normal Error (NE)
2	6 / 40	0 / 0	N/A	N/A	4 / 10	Interior	6,8,10,12	Interior Point Influence (IPI)
3	6 / 40	4 / 10	Axial	7,9,11,13	4 / 10	Interior	6,8,10,12	IPI
4	6 / 40	4 / 10	Axial	7,9,11,13	4 / 10	Exterior	6,8,10,12	High Leverage Outlier (HLO)
5	2 / 16	4 / 25	Axial	4,5,6,7	4 / 25	Interior	6,7,8,9	IPI
6	6 / 40	8 / 20	Axial	7,9,11,13	10 / 25	Interior	(8,9,10,11,12)*2	IPI, High Outlier Density (HOD)
7	9 / 80	16 / 20	Axial	10,12,14,16	20 / 25	Interior	(12,13,14,15,16)*4	IPI, HOD
8	2 / 16	4 / 25	Cloud - Diag	≅ (6,-6)*4	3 / 19	Exterior Cloud	(6,8,10) w/ replace.	HLO, HOD
9	2 / 16	4 / 25	Cloud - Off Line	≅ (6,-3)*4	3 / 19	Exterior Cloud	(6,8,10) w/ replace.	HLO, HOD
10	2 / 16	4 / 25	Cloud - Axial	≅ (5,0)*4	3 / 19	Exterior Cloud	(6,8,10) w/ replace.	HLO, HOD
11	2 / 16	4 / 25	Axial	4,5,6,7	3 / 19	Exterior	7,8,9	HLO, HOD
12	6 / 40	8 / 20	Axial	(7,9,11,13)*2	4 / 10	Exterior	6,8,10,12	HLO
13	9 / 80	16 / 20	Axial	(10,12,14,16)*4	8 / 10	Exterior	(12,14,16,18)*2	HLO, HOD
14	6 / 40	8 / 20	Axial	(4,5,6,7)*2	4 / 10	Exterior	6,8,10,12	HLO
15	6 / 40	8 / 20	Axial	(4,6,14,16)*2	6 / 15	Exterior	(5,10,15)*2	HLO
16	6 / 40	8 / 20	Axial	(7,9,11,13)*2	8 / 20	Exterior	(6,8,10,12)*2	HLO, HOD

Results from DS3 can be directly compared to those from DS2 to determine any impact that "good" high leverage points have on estimation accuracy. DS4 contains the same 10% high leverage points used in DS3 and moves the 10% outliers to the high leverage points, so that the high leverage outlier characteristic is modeled.

Datasets 5 through 7 focus on a high percentage of interior point outliers (25%), and only modify the number of model parameters across the three datasets. DS5 has two independent variables, DS6 has six, and DS7 has nine. Two of the three datasets (DS6 and DS7) are considered high outlier density because the percentage of outliers exceed the GM-estimate breakdown point of $1/p$. The larger dimension models should pose even larger challenges to GM-estimators and the zero breakdown point estimators.

Datasets 8 through 10 are the multiple point cloud configurations. Because small dimension models are easier to visualize, two-variable models are used. The first cloud dataset (DS8) consists of a cloud located along one of the diagonals of the cube design. This location effectively confuses the MVE estimate of leverage, causing the MVE algorithm to select the off-diagonal corner points as the high leverage points. The Krasker-Welsch (KW) weights method performs slightly better because although the cloud is not distinguishably identified as high leverage, the method does not assign extremely large weights to the off-diagonal interior points.

The dataset labeled DS9 locates the cloud not in-line with any two cube corner points so that the MVE performs well in identifying the cloud as high leverage. The KW weights method, again influenced by the cloud, does not assign significantly higher weights to the cloud points. As a result, most of the observations receive similar weights.

The cloud points for DS10 are located along the cube axis at a distance sufficiently far from the corner points. Unfortunately, MVE distances indicate the corner points furthest from the cloud are moderately high leverage points. The KW method is not largely influenced by the cloud so that the cloud points receive larger weights than the cube corner points. The errors for the outliers of all three of these datasets are selected in a similar fashion. Each dataset contains three outliers and each outlier is assigned a sign and magnitude (6, 8, or 10) randomly. The magnitudes are assigned with replacement, meaning that a magnitude may be selected more than once or not at all.

The remaining datasets (11 through 16) all place outliers on the high leverage points. The number of parameters and percentage of outliers are varied along with the location of the high leverage points. Each dataset contains 20% high leverage points located in different directions on the axes of the cube, but their distance from the cube center is varied. DS11 is a two-variable problem with axial leverage points all located about the same distance in all four directions from the center. Three of the four axial points are assigned high error values. The outlier positions are selected randomly from the four axial points and assigned one of three error values, each of similar magnitude. The outlier density is not high enough to be considered HOD, but this and all remaining datasets are high leverage outliers.

DS12 and DS13 are six and nine parameter datasets with 10% high leverage outliers. Again the high leverage points are located on different axes and the outliers among those points are randomly selected. DS14 is a slight modification of DS12, with the leverage points moved in towards the cube. These leverage points are near the inlier/outlier border as defined by the MVE distance cutoffs. DS15 modifies DS14 by moving two of the borderline X-space leverage points to locations between two and three times farther away from the cube center. The resulting design contains four moderate and four very high leverage points. The percentage of outliers is also increased from 10% to 15%, so six of the eight leverage points are selected at random to receive high error magnitudes. The final dataset (DS16) is another six parameter model that contains eight high leverage points, all of which are treated as outliers. The outlier density (20%) is considered high for this model dimension.

Run-to-run variation for all datasets is composed of differences in the values of random normal variates for the "good" points and differences in the signs of the outliers. Run-to-run variation produces datasets that differ in terms of estimation difficulty for least squares and the various robust techniques. For example, datasets with outliers all in the same direction (all the same sign) are more difficult for some techniques (including least squares) to accurately estimate. Fifty run variations are developed and model estimates are generated for all techniques for each run. Mean square errors of estimation are calculated for each run and averaged across all fifty runs. Analysis of previous experiments shows that the variances of the resulting average mean square errors of estimation are acceptably small. The correlation between technique estimates for a particular run variation is also high for most pair-wise technique comparisons. Therefore, the additional information of the correlation between technique estimates increases our confidence in the relative differences between technique AMSEE values.

5.3.3 Measuring Technique Performance

The experiment contains 16 technique/dataset (treatment) combinations and 11 robust techniques with 50 replicates per treatment combination, resulting in 8800 model estimations. Some estimations, such as least squares, are quick computationally, while others such as LTS and *S*-estimation take considerably longer, especially for the nine-variable, 80 observation problems. The estimated coefficients are compared to the true model coefficient by computing a mean square error of estimation (MSEE). Small MSEE values are desirable. The replicates within a treatment combination are used to calculate an average MSEE (AMSEE) using the expression

$$\text{AMSEE} = \text{mean}\left[\left(\hat{\beta}_R - \beta\right)'\left(\hat{\beta}_R - \beta\right)\right] \tag{5.10}$$

where $\hat{\beta}_R$ refers to the robust technique estimated coefficients and β is the vector of true model coefficients. Another performance statistic related to the AMSEE is the average mean square inefficiency ratio (AMSIR) which is a ratio of a regression technique's AMSEE to the least squares AMSEE.

$$\text{AMSIR} = \frac{\text{mean}\left[(\hat{\beta}_R - \beta)'(\hat{\beta}_R - \beta)\right]}{\text{mean}\left[(\hat{\beta}_{LS} - \beta)'(\hat{\beta}_{LS} - \beta)\right]} \tag{5.11}$$

The AMSIR represents the percent improvement (or in some instances degradation) over least squares in accuracy of estimation for a particular technique / dataset combination.

Statistics can be compiled based on the AMSEE values to further evaluate and compare the robust techniques and to help determine which techniques perform well over a variety of scenarios. The rank of a technique for an experiment run is determined by ordering the AMSEE values of the robust techniques from smallest to largest. The smallest AMSEE receives a rank of one. The ranks can then be summed to determine the best techniques (lowest overall rank). The top techniques can then be compared in terms of the standard deviation of the rank. Smaller rank standard deviations are desired because they indicate less weak areas.

Another statistic used to evaluate AMSEE performance is the percent over the minimum AMSEE for a particular scenario. Each technique AMSEE is divided by the top ranking technique AMSEE for each run. This percent over AMSEE can then be summed and techniques can be compared using this summed statistic. The added value of this statistic over the rank sum is that the spread in AMSEE values within runs is being measured. If one technique ranks second to another on three runs and is first on the fourth, it will have a larger rank sum. However, if it finished second by only 1% in AMSEE on each of the first three runs and is first on the fourth run by 10%, it will have a lower percent over AMSEE. Both statistics provide useful information and will be included in the technique comparison summary tables.

5.3.4 Experiment 1 Results

5.3.4.1 Robust Techniques versus Dataset Characteristics

The first technique performance study consists of observing the behavior of least squares and ten robust methods against a variety of outlier conditions. The datasets described in the previous section represent a mix of small and large models, interior and exterior **X**-space outliers, and moderate (10%) and high (20%) outlier densities. Discussion of technique performance will consist of general comments regarding classes of techniques versus various dataset types and specific comments covering each technique's performance in detail. Each technique's performance is summarized in Table 5.5. Results are reported in terms of AMSEE, relative robust technique rank, and percent over the minimum AMSEE for a particular dataset. Experiment totals are printed in bold below each statistic category.

5.3.4.2 General Comments

Techniques in the same class (with similar desirable properties) tend to perform in a similar fashion. Most theoretical findings regarding technique behavior were verified empirically in this study. For instance,

- No robust technique outperforms least squares for normally distributed error datasets

- The more efficient techniques (*M*-, MM-, and GM-estimation) perform better than the low efficiency techniques (LTS, and *S*-estimation) with normal data
- High efficiency techniques without bounded influence (*M*-estimation and most B-robust) perform superbly on interior **X**-space outliers, regardless of outlier density. The same techniques perform miserably on exterior **X**-space outliers, again regardless of outlier density
- Multiple stage techniques perform better than single stage techniques. This result may be simply due to the fact that each multiple stage technique tested has two or more desirable properties, while the single stage estimators only have a single desirable property
- No robust technique consistently outperforms all others. The two best techniques have an average ranking of 3.6 out of the ten robust methods
- No robust techniques excel in performance on all of the multiple point cloud datasets. None of the leverage measures employed is able to correctly identify all three cloud locations as containing high leverage points. Three robust techniques actually perform worse than least squares on all three cloud configurations

5.3.4.3 Dataset Performance Comparisons

- *Leverage (DS2) points versus no leverage (DS3) points*: All except LTS and *S*-estimation perform well. All techniques improve their performance when "good" leverage points are added. Least squares and most B-robust estimation AMSEE values are cut in half. All other techniques experience significant reductions, except *S*-estimation
- *Interior **X**-space outliers (DS3) versus Exterior **X**-space outliers (DS4):* *M*-estimation goes from best to worst and most B-robust also performs significantly worse with high leverage outliers. GM-estimation is only slightly impacted by the move to high leverage outliers. The

high breakdown point only techniques (LTS and *S*-estimation) perform significantly worse than the other robust methods

- *High Outlier Density (25%) Interior* ***X****-space outliers: 2 variable (DS5) versus 6 variable (DS6) versus 9 variable (DS7)*: Some of the theory regarding high breakdown is not necessarily supported in terms of technique performance on these datasets. A method with high breakdown properties (LTS) actually performed worse than the 0% breakdown *M*-estimator on all three sized models

Table 5.5. Experiment 1 - Outlier Location / Density Performance Results

AMSEE												
DS	Description	LS	M	MB-R	LTS	S	MM	GMCH	GMMZ	GMNP1	GMNP2	GMNP3
1	6V, 0%, No Lev	0.19	0.20	0.28	0.66	0.60	0.20	0.27	0.27	0.20	0.20	0.24
2	6V, 10%, Int, No Lev	1.90	0.22	0.40	0.73	0.50	0.22	0.35	0.38	0.32	0.33	0.32
3	6V, 10%, Int, Lev	0.93	0.14	0.22	0.57	0.41	0.14	0.23	0.23	0.20	0.21	0.22
4	6V, 10%, Ext, Lev	1.82	1.84	1.40	0.73	0.56	0.31	0.36	0.30	0.28	0.25	0.29
5	2V, 25%, Int, Lev	1.17	0.24	0.19	0.26	0.13	0.30	0.27	0.33	0.44	0.59	0.31
6	6V, 25%, Int, Lev	0.97	0.09	0.16	0.32	0.18	0.10	0.18	0.21	0.29	0.39	0.19
7	9V, 25%, Int, Lev	0.85	0.05	0.10	0.29	0.12	0.05	0.18	0.13	0.23	0.27	0.14
8	2V, 19%, Cloud, Lev	0.65	0.95	0.93	0.64	0.44	0.43	0.51	0.74	0.71	0.62	0.35
9	2V, 19%, Cloud, Lev	0.84	1.31	1.32	0.82	0.52	0.52	0.49	0.84	0.94	0.76	0.37
10	2V, 19%, Cloud, Lev	1.32	1.72	1.58	0.57	0.48	0.78	0.61	0.87	1.22	1.04	0.47
11	2V, 19%, Ext, Lev	1.97	1.48	1.25	0.77	0.62	0.94	0.75	0.48	0.56	0.52	0.56
12	6V, 10%, Ext, Lev	1.62	0.87	0.92	0.61	0.46	0.30	0.35	0.26	0.24	0.23	0.27
13	9V, 10%, Ext, Lev	2.82	0.44	0.68	0.44	0.26	0.10	0.17	0.18	0.17	0.15	0.15
14	6V, 10%, Ext, Lev	2.07	0.23	0.47	0.59	0.45	0.22	0.33	0.33	0.27	0.27	0.28
15	6V, 15%, Ext, Lev	2.86	1.38	1.39	0.62	0.48	0.38	0.39	0.34	0.36	0.33	0.35
16	6V, 20%, Ext, Lev	2.79	3.11	3.14	1.25	0.97	1.29	0.83	0.44	0.57	0.48	0.72
	Sum	**24.77**	**14.28**	**14.43**	**9.85**	**7.18**	**6.29**	**6.26**	**6.33**	**7.01**	**6.64**	**5.22**

Robust Technique Ranking											
DS	Description	M	MB-R	LTS	S	MM	GMCH	GMMZ	GMNP1	GMNP2	GMNP3
1	2V, 12%, Int., No Lev	4	8	10	9	1	6	7	3	2	5
2	6V, 10%, Int, No Lev	2	8	10	9	1	6	7	4	5	3
3	10V, 10%, Int., No Lev	1	6	10	9	2	8	7	3	4	5
4	2V, 19%, Int., No Lev	10	9	8	7	5	6	4	2	1	3
5	6V, 20%, Int, No Lev	3	2	4	1	6	5	8	9	10	7
6	10V, 20%, Int., No Lev	1	3	9	4	2	5	7	8	10	6
7	2V, 12%, Int., Lev	1	3	10	4	2	7	5	8	9	6
8	6V, 10%, Int, Lev	10	9	6	3	2	4	8	7	5	1
9	10V, 10%, Int., Lev	9	10	6	3	4	2	7	8	5	1
10	2V, 19%, Int., Lev	10	9	3	2	5	4	6	8	7	1
11	6V, 20%, Int, Lev	10	9	7	5	8	6	1	4	2	3
12	10V, 20%, Int., Lev	9	10	8	7	5	6	3	2	1	4
13	2V, 12%, Ext, Lev	8	10	9	7	1	5	6	4	3	2
14	6V, 10%, Ext, Lev	2	9	10	8	1	6	7	4	3	5
15	10V, 10%, Ext, Lev	9	10	8	7	5	6	2	4	1	3
16	2V, 19%, Ext, Lev	9	10	7	6	8	5	1	3	2	4
	Sum of Ranks	**98**	**125**	**125**	**91**	**58**	**87**	**86**	**81**	**70**	**59**
	Std Dev of Ranks	**3.9**	**2.8**	**2.2**	**2.6**	**2.4**	**1.4**	**2.4**	**2.5**	**3.1**	**1.9**

Table 5.5. Cont.

Percent Over Minimum AMSEE

DS	Description	M	MB-R	LTS	S	MM	GMCH	GMMZ	GMNP1	GMNP2	GMNP3
1	6V, 0%, No Lev	2%	42%	228%	199%	0%	34%	35%	2%	0%	20%
2	6V, 10%, Int, No Lev	0%	85%	237%	128%	0%	59%	76%	48%	50%	45%
3	6V, 10%, Int, Lev	0%	55%	307%	194%	3%	65%	62%	46%	48%	53%
4	6V, 10%, Ext, Lev	632%	456%	190%	124%	24%	42%	19%	10%	0%	16%
5	2V, 25%, Int, Lev	84%	42%	96%	0%	128%	103%	154%	236%	349%	136%
6	6V, 25%, Int, Lev	0%	73%	255%	95%	9%	102%	130%	221%	328%	108%
7	9V, 25%, Int, Lev	0%	110%	485%	148%	1%	264%	172%	357%	442%	181%
8	2V, 19%, Cloud, Lev	167%	162%	80%	24%	22%	42%	107%	101%	76%	0%
9	2V, 19%, Cloud, Lev	252%	256%	120%	40%	41%	32%	126%	155%	104%	0%
10	2V, 19%, Cloud, Lev	265%	236%	20%	2%	65%	30%	86%	159%	122%	0%
11	2V, 19%, Ext, Lev	212%	163%	61%	30%	98%	59%	0%	17%	10%	17%
12	6V, 10%, Ext, Lev	288%	307%	170%	103%	32%	54%	14%	6%	0%	20%
13	9V, 10%, Ext, Lev	341%	581%	345%	162%	0%	67%	79%	66%	50%	46%
14	6V, 10%, Ext, Lev	5%	117%	170%	108%	0%	52%	53%	26%	25%	28%
15	6V, 15%, Ext, Lev	318%	320%	88%	46%	14%	19%	3%	10%	0%	6%
16	6V, 20%, Ext, Lev	608%	614%	183%	120%	194%	90%	0%	29%	10%	63%
	Sum	**3174%**	**3619%**	**3033%**	**1522%**	**632%**	**1113%**	**1115%**	**1489%**	**1613%**	**738%**

AMSIR

DS	Description	M	MB-R	LTS	S	MM	GMCH	GMMZ	GMNP1	GMNP2	GMNP3
1	6V, 0%, No Lev	1.07	1.48	3.43	3.13	1.05	1.40	1.41	1.07	1.05	1.25
2	6V, 10%, Int, No Lev	0.11	0.21	0.39	0.26	0.11	0.18	0.20	0.17	0.17	0.17
3	6V, 10%, Int, Lev	0.15	0.24	0.62	0.45	0.16	0.25	0.25	0.22	0.22	0.23
4	6V, 10%, Ext, Lev	1.01	0.77	0.40	0.31	0.17	0.20	0.16	0.15	0.14	0.16
5	2V, 25%, Int, Lev	0.21	0.16	0.22	0.11	0.26	0.23	0.29	0.38	0.51	0.27
6	6V, 25%, Int, Lev	0.09	0.16	0.33	0.18	0.10	0.19	0.22	0.30	0.40	0.20
7	9V, 25%, Int, Lev	0.06	0.12	0.34	0.14	0.06	0.21	0.16	0.27	0.31	0.16
8	2V, 19%, Cloud, Lev	1.47	1.44	0.99	0.68	0.67	0.78	1.14	1.10	0.96	0.55
9	2V, 19%, Cloud, Lev	1.55	1.56	0.97	0.61	0.62	0.58	0.99	1.12	0.90	0.44
10	2V, 19%, Cloud, Lev	1.31	1.20	0.43	0.36	0.59	0.46	0.66	0.93	0.79	0.36
11	2V, 19%, Ext, Lev	0.75	0.64	0.39	0.31	0.48	0.38	0.24	0.28	0.27	0.28
12	6V, 10%, Ext, Lev	0.54	0.57	0.38	0.28	0.18	0.21	0.16	0.15	0.14	0.17
13	9V, 10%, Ext, Lev	0.16	0.24	0.16	0.09	0.04	0.06	0.06	0.06	0.05	0.05
14	6V, 10%, Ext, Lev	0.11	0.23	0.28	0.22	0.10	0.16	0.16	0.13	0.13	0.13
15	6V, 15%, Ext, Lev	0.48	0.48	0.22	0.17	0.13	0.14	0.12	0.13	0.12	0.12
16	6V, 20%, Ext, Lev	1.12	1.13	0.45	0.35	0.46	0.30	0.16	0.20	0.17	0.26
	Sum	**10.18**	**10.62**	**9.98**	**7.66**	**5.19**	**5.74**	**6.38**	**6.65**	**6.33**	**4.79**

5.3.4.4 Experiment 1 Individual Technique Performance

M-estimation

This technique either performed at the top or near the bottom. The determinant of performance for this technique is the location of the outliers. The lack of a bounded influence objective function resulted in poor fits for models with high leverage point outliers, including the cloud datasets. *M*-estimation performed impressively on interior point outlier datasets. The use of the redescending ψ-function drives the outliers to zero weight in many instances, improving the error of estimation. Unfortunately, using this type of ψ-function can also lead to poor fits if the initial estimate is not good. There is one anomaly regarding *M*-estimation and outlier point location. DS14 contains exterior **X**-space outliers located near the interior/exterior border. As a result, the outliers are not located on significantly high leverage points. *M*-estimation is ranked 2nd among robust methods in this case. Thus *M*-estimation tends to perform well on datasets with outliers at moderate leverage point locations.

Most B-robust

This technique tends to perform poorly relative to the other robust techniques regardless of dataset configuration. It is important to note that although most B-robust is not an outstanding robust method, it is still a significant improvement over least squares. In terms of overall AMSEE, the most B-robust method is nearly twice as accurate as least squares (14.3 versus 24.8). This technique does perform well with high outlier densities in the interior **X**-space. However, when exterior outliers are present, this method performs similarly to *M*-estimation. Most B-robust is a weighted L_1-norm method with weights that are indicators of leverage. Monte Carlo studies confirm the published accounts that L_1-norm estimation is largely influenced by high leverage outliers when present. The weights used in the most B-robust technique do not appear to substantially improve estimation under high leverage outlier conditions.

LTS

The Least Trimmed Sums of Squares Technique ties with the most B-robust technique as the worst performing robust alternative. This method struggles with non-outlier datasets and with moderate outlier density conditions. Little improvement is observed in the high outlier density datasets as the technique is only average relative to the other estimators. A suggestion for future tests involving LTS is to modify the random subsample size to increase efficiency and sacrifice some degree of breakdown.

S-estimation

S-estimation is a high breakdown method that expectedly performs better as the outlier density increases. Its weakest showings involve datasets with no outliers and those with 10% outliers. However, for datasets with high outlier densities, including the cloud datasets, *S*-estimation is a solid performer (average rank = 3.5 for those eight datasets).

MM-estimation

MM-estimation is a two stage technique that inherits and maintains high breakdown from its initial *S*-estimate and is high efficiency due to its *M*-estimation final estimate convergence. Unfortunately, MM-estimation does not explicitly have bounded influence. This lack of explicit bounded influence results in estimation problems against some high leverage outlier situations involving small to moderate model dimensions and high outlier density (DS11 and DS16). When

20-25% of a dataset are high leverage point outliers, MM-estimation ranks eighth out of the ten robust techniques. However, MM-estimation performs well in general, even in most situations involving high leverage outliers. This technique has high efficiency, performs near or at the top against interior point outliers, and performs well against multiple point clouds and moderate density exterior point outliers. MM-estimation is one of the two overall best performing robust techniques.

GM (Coakley and Hettmansperger)

This multiple stage estimator starts with LTS initially and performs only one Newton iteration towards convergence using a GM objective function. Although the authors show this technique has all three desirable properties, it only performs average relative to the other robust methods. Comparing its results to LTS shows that the GM objective improves the accuracy of estimation for nearly all scenarios. Unfortunately, LTS is not the best single stage robust technique, so in some cases the initial estimate requires substantial improvement. Tests on variations of this method using fully iterative GM convergence show that the one-step method performs better. Overall, this one-step technique does not have any serious weaknesses as demonstrated by the technique yielding the second smallest overall AMSEE.

GM (Marazzi)

The Marazzi method is a GM method using most B-robust estimation initially, followed by a GM objective consisting of Krasker-Welsch estimates of leverage, and the Huber ψ-function. Initial tests with other GM techniques indicate that this technique performs well under most conditions. This study concludes that, relative to other robust techniques, the Marazzi method is an above average performer. It performs below average in efficiency and in treating interior point outliers, regardless of density. It performs well against exterior point (high leverage) outliers, particularly high density outlier datasets. This method's main drawback is that it tends to downweight too many observations, including non-outliers. This characteristic is likely the cause of its below average efficiency, and poor performance against moderate outlier density problems and multiple point clouds.

GM (New Proposal - 1)

This technique is the first of the new GM-estimation proposals that performed well enough in Chapter 4 to be considered for overall robust technique performance consideration. This method is a modification of the Marazzi technique which attempts to moderate the severe downweighting by scaling the Krasker-Welsch weights. GMNP1 performs as well overall as the Marazzi approach. The modification achieves the intended results improving accuracy in efficiency and in moderate outlier situations (10% outlier density). However, performance is slightly degraded for high outlier density problems.

GM (New Proposal - 2)

The purpose of the second GM-estimation proposal is to replace the initial most B-robust estimate with the high breakdown, better performing *S*-estimate. The intent of the replacement is to improve overall estimation, particularly in the high outlier density problems. This technique performs especially well in terms of efficiency and against high leverage outliers (average rank = 1.86 for seven datasets) not including the multiple point clouds. Poor performance is evident, however, in high density, interior point outlier datasets (average rank = 9.7 for three datasets) and the multiple point cloud datasets. This technique shows promising results with some areas indicating a need for technique refinement.

GM (New Proposal - 3)

This third alternative GM proposal (GMNP3) further refines the two previous alternatives by restricting the convergence algorithm to one IRLS iteration. The one-step convergence technique simplifies the estimation process, which consists of an initial *S*-estimate followed by a single weighted least squares iteration using a GM objective function. The measures of leverage and ψ-function are unchanged from GMNP1. This alternative is comparable to MM-estimation in overall performance. It has the lowest overall AMSEE of 5.2, which is 16% lower than the next best robust technique. In terms of overall rank, it is essentially equal to MM-estimation. GMNP3 improves on GMNP2 by taking large initial steps away from the initial *S*-estimate in the correct direction and small initial steps in the incorrect direction (towards a worse estimate). Large increases in estimation accuracy are achieved in the multiple point cloud problem (top rank for all three datasets), and the high density interior point outlier problems. Slight decreases in efficiency and exterior point outlier fit are observed. In general, this technique has no serious weaknesses.

5.4 Experiment 2 - Structured Designs

5.4.1 Experiment Description

The second experiment is a natural extension of the first experiment. The goal is to develop a balanced and comprehensive set of scenarios that robust techniques may encounter in empirical situations. The intent of this sequential design process is to increase the exposure of the previously tested robust techniques to a greater number of outlier scenarios. More specifically, the purpose of the second experiment is to:

- Investigate the efficiency performance in more detail

- Increase the level of difficulty on the multiple cloud problems by ensuring that all outliers are located in the same direction
- Introduce a group of datasets with extremely large outliers. The magnitude of these outliers is designed to simulate conditions of typing or coding errors involving decimal point shifts
- Establish an outlier-location / leverage-content factor so that the scenarios contain outlier locations and leverage points in all conceivable combinations. An additional goal is to place a mix of outliers in both the interior **X**-space points and the exterior **X**-space (high leverage) points for some datasets

Attempts to implement these four initiatives led to the development of four distinct sub-experiments for investigating: 1) efficiency, 2) the multiple point cloud situation, 3) large magnitude outliers, and 4) a complete mix of model dimension, outlier location, and outlier density. To avoid confusion in future discussions between the two sequential experiments and four sub-experiments within Experiment 2, we will continue to refer to the four designs of Experiment 2 as sub-experiments or sub-designs.

The purpose of conducting these sub-experiments is not necessarily to determine the most significant factors affecting the robust techniques. The purpose is primarily to determine which techniques perform the best overall when exposed to the most diverse and complete set of outlier scenarios. Thus, factor number and factor level consistency are not primary considerations. By trying to keep the experiment as simple and efficient as possible, only those levels of the factors of interest that have shown to be important in previous studies are included.

The efficiency sub-experiment is a two-factor mixed level design. The factors are the number of model parameters and leverage content. Three levels are used for the numbers of model coefficients; two, six and ten. The leverage factor has two levels; datasets with leverage points and those without leverage points. The hypothesis regarding the effect of leverage points on efficiency is that those techniques without bounded influence may gain more information by the addition of "good" high leverage points. Obviously, this characteristic can lead to poor estimation of the majority of the data if the high leverage points are not in-line with the bulk of the data. Another apriori hypothesis is that increasing the number of model parameters will compound the difficulties of low efficiency estimators in accurately estimating normal error datasets. Table 5.6 below shows the experimental treatment combinations for the efficiency sub-design.

Table 5.6. Sub-Experiment 2.1 - Efficiency Test

Number of Independent Variables	Leverage Content
2	No Leverage
6	No Leverage
10	No Leverage
2	20% High Leverage Points
6	20% High Leverage Points
10	20% High Leverage Points

Sub-design 2.2 involves a test of robust technique performance against datasets containing multiple point clouds. Multiple point clouds are clusters of two or more observations located a significant distance away in the **X**-space from the majority of the observations. Experimental runs for this sub-design are developed similar to the cloud datasets from the first experiment (DS8, 9, and 10). Errors for the cloud point outliers in the first experiment had random sign, meaning the +/- direction for the outliers is randomly chosen. This approach resulted in the majority of the replicates having outliers of mixed sign. This type of outlier configuration caused some estimators, including least squares, to fit between the positive and negative signed outliers, which is not far from the true line.

To more aggressively challenge the techniques using the multiple point clouds, the outliers in sub-design 2.2 all have a positive sign, causing estimators that do not effectively downweight outlier impact to develop model fits detectably different from the true regression plane. Table 5.7 shows the approximate (X1, X2) location of the 4-point clouds used in each experimental run. The actual points are no more than 5% away from the location specified in the table.

Table 5.7. Sub-Experiment 2.2 - Multiple Point Cloud Test

Cloud Description	(X1, X2) General Locations
On Cube Diagonal	(6, -6)
Off-Line with Cube Points	(6, -3)
On Cube Center Axis	(5, 0)

A situation that often arises in practical experience is the case of outliers caused by typing or coding errors. Not only are the actual values incorrectly entered, but many times the decimal point is either missing when needed or shifted to an improper location. The result of such an error is a point that is an order of magnitude away from the rest of the data. The purpose of this sub-design is to determine performance degradations of estimators when large error values are present. Improper coding of response values is one cause of these large error values. Three dataset configurations are used to test the large error performance. The three configurations include a mix of number of model parameters, outlier density, outlier location, and high leverage point presence. These combinations, along with the magnitude of the errors are detailed in Table 5.8.

Table 5.8. **Sub-Experiment 2.3 - Large Error Outlier Test**

Dataset Description	Outlier Magnitude
2 Variable, 20% Interior Outliers, 20% Leverage Points	10, 20, 30
6 Variable, 20% Exterior Outliers, 20% Leverage Points	10, 20, 30, 40, 50, 60, 70, 80
10 Variable, 10% Interior Outliers, No Leverage Points	20, 30, 40, 50, 60, 70, 80, 90

The most extensive component of the robust technique simulation study is sub-experiment 2.4, designed to study the breakdown and bounded influence capabilities of these robust methods. Previous simulations have shown that three factors significantly affect technique performance; outlier location/leverage content, number of model parameters and outlier density. The location of the outliers and presence of leverage points combine to form a single factor in the arrangements shown in the first column of Table 5.9.

Table 5.9. **Sub-Experiment 2.4 - Outlier Location/Density and Model Dimension**

Outlier Location / Leverage Content	Number of Parameters	Outlier Density
Interior / No Leverage	2	10%
Interior / Leverage	6	20%
Exterior / Leverage	10	
Interior & Exterior / Leverage		

The two, six and ten variable levels should be considered more of a categorical factor than a quantitative factor. The two variable level represents a small regression model, the six variable level is representative of moderate sized regression models and the ten variable level represents large regression models. The third factor, outlier density is designed for two levels, 10% outliers and 20% outliers. The 10% level is a typical number of outliers for many empirical datasets (see Hampel, et. al. 1986). The 20% outlier density level represents a large number of discrepant observations and is used primarily to challenge the breakdown capabilities of robust estimators.

These three factors at four, three, and two levels respectively result in a 24-run experiment. All runs in each of the three component experiments are replicated 50 times so that an average mean square of estimation (AMSEE) statistic is calculated for each technique. The resulting AMSEE values are compared across techniques for each treatment combination. The 24 runs included in this sub-design are described in Table 5.10. The code for this simulation is provided in Appendix D.

Table 5.10. Experiment Runs for Outlier Location / Density and Leverage Presence Test

DS	Vars / Sample	Leverage # / %	Leverage Distance	Outliers # / %	Outlier Location in X-Space	Outlier Magnitude
1	2 / 16	0 / 0	N/A	2 / 12	Interior	6, 8
2	6 / 40	0 / 0	N/A	4 / 10	Interior	6, 8, 10, 12
3	10 / 80	0 / 0	N/A	8 / 10	Interior	(10, 12, 14, 16) * 2
4	2 / 16	0 / 0	N/A	3 / 19	Interior	6, 8, 10
5	6 / 40	0 / 0	N/A	8 / 20	Interior	(6, 8, 10, 12) * 2
6	10 / 80	0 / 0	N/A	16 / 20	Interior	(10, 12, 14, 16) * 4
7	2 / 16	4 / 25	4, 5, 6, 7	2 / 12	Interior	6, 8
8	6 / 40	8 / 20	7, 9, 11, 13	4 / 10	Interior	6, 8, 10, 12
9	10 / 80	16 / 20	10, 12, 14, 16	8 / 10	Interior	(10, 12, 14, 16) * 2
10	2 / 16	4 / 25	4, 5, 6, 7	3 / 19	Interior	6, 8, 10
11	6 / 40	8 / 20	7, 9, 11, 13	8 / 20	Interior	(6, 8, 10, 12) * 2
12	10 / 80	16 / 20	10, 12, 14, 16	16 / 20	Interior	(10, 12, 14, 16) * 4
13	2 / 16	4 / 25	4, 5, 6, 7	2 / 12	Exterior	6, 8
14	6 / 40	8 / 20	7, 9, 11, 13	4 / 10	Exterior	6, 8, 10, 12
15	10 / 80	16 / 20	10, 12, 14, 16	8 / 10	Exterior	(10, 12, 14, 16) * 2
16	2 / 16	4 / 25	4, 5, 6, 7	3 / 19	Exterior	6, 8, 10
17	6 / 40	8 / 20	7, 9, 11, 13	8 / 20	Exterior	(6, 8, 10, 12) * 2
18	10 / 80	16 / 20	10, 12, 14, 16	16 / 20	Exterior	(10, 12, 14, 16) * 4
19	2 / 16	4 / 25	4, 5, 6, 7	2 / 12	Int & Ext	6, 8
20	6 / 40	8 / 20	7, 9, 11, 13	4 / 10	Int & Ext	6, 8, 10, 12
21	10 / 80	16 / 20	10, 12, 14, 16	8 / 10	Int & Ext	(10, 12, 14, 16) * 2
22	2 / 16	4 / 25	4, 5, 6, 7	3 / 19	Int & Ext	6, 8, 10
23	6 / 40	8 / 20	7, 9, 11, 13	8 / 20	Int & Ext	(6, 8, 10, 12) * 2
24	10 / 80	16 / 20	10, 12, 14, 16	16 / 20	Int & Ext	(10, 12, 14, 16) * 4

This table details the leverage magnitudes, as well as the error magnitudes prescribed for each configuration. Leverage and error magnitudes are determined by the robust technique performances on previous outlier location and magnitude sensitivity studies. Technique performance tends to level off beyond a certain error value and increases in performance beyond that magnitude are not observed. Error magnitudes are adjusted slightly for models of different dimension. The resulting magnitudes produce datasets that fully challenge the robust techniques.

5.4.2 Introduction of New GM-Estimation Techniques

The results of the first experiment indicate that perhaps some improvements can be made in the best techniques to strengthen their weak areas and hopefully not degrade their strengths. For MM-estimation, one suggestion is to add a measure of leverage which may enhance its estimation accuracy against high leverage outliers. For GMNP3, there may be a modification that will improve its estimation against interior point outliers. These desires led to the development of two new GM-estimators, GMNP4 and GMNP5.

GMNP4 is a proposed enhancement of MM-estimation, which starts with an initial MM-estimate and performs one weighted least squares step using a GM objective function with Krasker-Welsch measures of leverage and the Huber ψ-function. Unfortunately, this estimator moved more in the wrong direction than it moved in an improved direction. Overall, MM-estimation performed better, so the alternative GMNP4 was not included in future tests.

GMNP5 is a slight modification of GMNP3, using the knowledge of the components of MM-estimators. In MM-estimation, both the χ-function of the initial estimate and the ψ-function of the final *M*-estimate use the hard redescending Tukey biweight formula. The suggestion is to change the ψ-function in the GM objective of GMNP3 to Tukey's biweight and observe the results.

Comparisons of the Huber versus Tukey ψ-function have been performed previously on fully iterated GM-estimators. The results of these comparisons indicate that, in some instances, the fully iterated hard redescending function results in final estimates distinctly different from the true model parameters. However, using only one iteration may avoid these potential problems. The initial test results using GMNP5 are encouraging so it is included in the following experiment. The alternative GMNP1 was dropped from consideration due to its lackluster performance in the first experiment.

5.4.3 Experiment 2 Implementation

The entire second experiment contains 36 dataset configurations among the four sub-experiments. Run times for each configuration are largely a function of the number of parameters and sample size. One replicate of a design point requires eleven model estimations. For the large problems (10 variables and 80 sample observations), a replicate requires nearly a minute of processing time on an 486-DX2 66 PC. Thus, a full run for these large problems can take nearly an hour for eleven estimations of all fifty replicates. The small problems take only five seconds per replicate and about four minutes per run. The slow run times of large problems is due to the poor convergence rates of the high breakdown estimators. Random subsample approximations are used for the two high breakdown methods, LTS and S-estimation. To reduce the within replicate estimation variability, estimates of those methods used as single stage techniques *and* initial estimates in multiple stage methods are made only once per replicate. This approach, applied to the high breakdown methods, also considerably reduces overall run time.

5.4.4 Experiment 2 General Performance Comments

Comments regarding overall robust technique performance on types of datasets are provided in this section. Also discussed are the sub-experiment general findings and the overall Experiment 2 performance results. Individual technique performance is discussed separately in the following section. Strengths and weaknesses of each method will be discussed relative to each sub-design. The specific results are provided in Table 5.11.

5.4.4.1 Efficiency Test

- Increasing the number of parameters decreases the AMSEE (increases the accuracy of estimation) for all techniques, including least squares
- Adding good leverage points increases the accuracy of estimation for all techniques, although some increase more than others. In general, the GM-estimation methods do not improve as much when good leverage points are added
- Overall, three techniques have superior efficiency (*M*-estimation, MM-estimation, and GMNP2), five techniques have high efficiency (most B-robust, GMCH, GMMZ, GMNP3, and GMNP5), and two methods have low efficiency (LTS and *S*-estimation). The associated AMSIR values for superior, high and low efficiency techniques are about 1.3, 2.0 and 5.0. The superior techniques have AMSEE values about 30% higher than least squares, the high efficiency techniques are twice as high as least squares and the low efficiency techniques are five times less accurate than least squares.

Table 5.11. Experiment 2 - Efficiency, Point Cloud, and Large Magnitude Outlier Tests

AMSEE

	DS	LS	M	MB-R	LTS	S	MM	GMCH	GMMZ	GMNP2	GMNP3	GMNP5
Efficiency	2V, No Lev	0.20	0.22	0.33	0.74	0.55	0.21	0.28	0.30	0.21	0.24	0.25
	6V, No Lev	0.19	0.21	0.27	0.80	0.69	0.21	0.31	0.26	0.20	0.25	0.29
	10V, No Lev	0.13	0.16	0.22	0.60	0.53	0.15	0.21	0.21	0.15	0.18	0.20
	2V, Lev	0.09	0.11	0.13	0.53	0.38	0.12	0.15	0.17	0.12	0.15	0.22
	6V, Lev	0.07	0.08	0.11	0.47	0.38	0.10	0.16	0.17	0.11	0.13	0.19
	10V, Lev	0.04	0.05	0.06	0.34	0.28	0.06	0.12	0.10	0.07	0.09	0.13
	Sum	**0.71**	**0.82**	**1.11**	**3.48**	**2.81**	**0.85**	**1.23**	**1.20**	**0.85**	**1.04**	**1.29**
Multiple Point Cloud	On Diag	2.49	3.62	3.26	0.43	0.25	0.18	0.50	1.88	2.68	0.36	0.18
	Off Diag	2.41	3.57	3.53	0.62	0.42	0.30	0.51	1.35	2.05	0.43	0.29
	On Axis	2.16	1.80	2.04	0.51	0.38	0.41	0.60	0.95	1.17	0.41	0.23
	Sum	**7.06**	**8.98**	**8.83**	**1.56**	**1.05**	**0.90**	**1.62**	**4.18**	**5.90**	**1.20**	**0.70**
Large Magnitude Outlier	2V,20%,Int,Lev	7.57	0.13	0.24	0.29	0.19	0.12	0.23	0.25	0.35	0.26	0.16
	6V,20%,Ext,Lev	84.87	38.69	25.03	0.71	0.42	0.26	0.48	0.45	0.45	0.43	0.25
	10V,10%,Int,No Lev	48.77	0.14	0.28	0.64	0.38	0.14	0.24	0.27	0.22	0.22	0.15
	Sum	**141.21**	**38.96**	**25.55**	**1.63**	**0.99**	**0.52**	**0.94**	**0.97**	**1.02**	**0.91**	**0.56**
	Weighted Mean AMSEE	**16.51**	**5.37**	**3.88**	**0.55**	**0.38**	**0.20**	**0.35**	**0.64**	**0.82**	**0.29**	**0.21**

Robust Technique Rank

	DS		M	MB-R	LTS	S	MM	GMCH	GMMZ	GMNP2	GMNP3	GMNP5
Efficiency	2V, No Lev		3	8	10	9	1	6	7	2	4	5
	6V, No Lev		3	6	10	9	2	8	5	1	4	7
	10V, No Lev		3	8	10	9	2	7	6	1	4	5
	2V, Lev		1	4	10	9	3	5	7	2	6	8
	6V, Lev		1	4	10	9	2	6	7	3	5	8
	10V, Lev		1	3	10	9	2	7	6	4	5	8
	Sum		**12**	**33**	**60**	**54**	**12**	**39**	**38**	**13**	**28**	**41**
Multiple Point Cloud	On Diag		10	9	5	3	1	6	7	8	4	2
	Off Diag		10	9	6	3	2	5	7	8	4	1
	On Axis		9	10	5	2	4	6	7	8	3	1
	Sum		**29**	**28**	**16**	**8**	**7**	**17**	**21**	**24**	**11**	**4**
Large Magnitude Outlier	2V,20%,Int,Lev		2	6	9	4	1	5	7	10	8	3
	6V,20%,Ext,Lev		10	9	8	3	2	7	5	6	4	1
	10V,10%,Int,No Lev		2	8	10	9	1	6	7	4	5	3
	Sum		**14**	**23**	**27**	**16**	**4**	**18**	**19**	**20**	**17**	**7**
	Weighted Mean Rank		**5.44**	**7.50**	**8.11**	**5.67**	**1.89**	**6.06**	**6.56**	**5.61**	**4.67**	**3.50**

Table 5.11. cont.

Percent Over Minimum AMSEE

	DS		M	MB-R	LTS	S	MM	GMCH	GMMZ	GMNP2	GMNP3	GMNP5
Efficiency	2V, No Lev		3%	58%	254%	163%	0%	33%	42%	0%	12%	18%
	6V, No Lev		6%	32%	297%	241%	2%	55%	28%	0%	23%	44%
	10V, No Lev		5%	46%	302%	255%	2%	41%	38%	0%	20%	37%
	2V, Lev		0%	22%	390%	250%	12%	38%	61%	10%	39%	103%
	6V, Lev		0%	40%	515%	403%	28%	106%	117%	39%	76%	152%
	10V, Lev		0%	28%	602%	470%	22%	151%	102%	41%	84%	174%
	Sum		**15%**	**225%**	**2360%**	**1782%**	**66%**	**423%**	**388%**	**90%**	**254%**	**527%**
Multiple Point Cloud	On Diag		1887%	1692%	136%	38%	0%	177%	933%	1374%	97%	1%
	Off Diag		1145%	1131%	117%	47%	4%	79%	372%	615%	51%	0%
	On Axis		672%	778%	119%	63%	78%	159%	308%	402%	74%	0%
	Sum		**3705%**	**3601%**	**372%**	**149%**	**83%**	**415%**	**1613%**	**2391%**	**223%**	**1%**
Large Magnitude Outlier	2V,20%,Int,Lev		6%	101%	142%	64%	0%	90%	110%	197%	116%	32%
	6V,20%,Ext,Lev		15324%	9879%	183%	66%	2%	91%	80%	81%	72%	0%
	10V,10%,Int,No Lev		1%	100%	353%	172%	0%	70%	89%	56%	57%	10%
	Sum		**15331%**	**10080%**	**677%**	**301%**	**2%**	**251%**	**279%**	**333%**	**245%**	**42%**
	Weighted Mean AMSEE		**2116%**	**1533%**	**248%**	**149%**	**13%**	**98%**	**232%**	**308%**	**66%**	**34%**

AMSIR

	DS		M	MB-R	LTS	S	MM	GMCH	GMMZ	GMNP2	GMNP3	GMNP5
Efficiency	2V, No Lev		1.10	1.69	3.79	2.82	1.07	1.42	1.52	1.08	1.20	1.27
	6V, No Lev		1.15	1.43	4.30	3.70	1.11	1.68	1.38	1.08	1.33	1.56
	10V, No Lev		1.18	1.63	4.50	3.98	1.15	1.58	1.55	1.12	1.34	1.53
	2V, Lev		1.20	1.46	5.89	4.20	1.35	1.66	1.94	1.32	1.67	2.44
	6V, Lev		1.12	1.57	6.92	5.66	1.43	2.31	2.44	1.56	1.98	2.83
	10V, Lev		1.18	1.51	8.28	6.72	1.43	2.95	2.38	1.66	2.17	3.23
	Sum		**6.94**	**9.30**	**33.68**	**27.08**	**7.54**	**11.61**	**11.21**	**7.82**	**9.70**	**12.86**
Multiple Point Cloud	On Diag		1.45	1.31	0.17	0.10	0.07	0.20	0.75	1.07	0.14	0.07
	Off Diag		1.48	1.46	0.26	0.18	0.12	0.21	0.56	0.85	0.18	0.12
	On Axis		0.83	0.94	0.24	0.18	0.19	0.28	0.44	0.54	0.19	0.11
	Sum		**3.76**	**3.72**	**0.67**	**0.45**	**0.39**	**0.69**	**1.75**	**2.47**	**0.51**	**0.30**
Large Magnitude Outlier	2V,20%,Int,Lev		0.02	0.03	0.04	0.03	0.02	0.03	0.03	0.05	0.03	0.02
	6V,20%,Ext,Lev		0.46	0.29	0.01	0.00	0.00	0.01	0.01	0.01	0.01	0.00
	10V,10%,Int,No Lev		0.00	0.01	0.01	0.01	0.00	0.00	0.01	0.00	0.00	0.00
	Sum		**0.48**	**0.33**	**0.06**	**0.04**	**0.02**	**0.04**	**0.04**	**0.06**	**0.04**	**0.03**
	Weighted Mean Rank		**0.86**	**0.97**	**1.95**	**1.56**	**0.46**	**0.73**	**0.82**	**0.71**	**0.60**	**0.75**

5.4.4.2 Multiple Point Cloud

Recall that modifications to the first experiment are made to challenge the techniques more aggressively. The errors are all pointed in the same direction and the magnitude of the outliers in the cloud is increased. The changes increases the AMSEE of least squares and of the robust methods which have a difficult time with this type of problem. Further examination of the two experiments reveals that non-bounded influence techniques such as MM-estimation, which does not benefit from explicit leverage measures, estimates much better due to the increased magnitude of the outliers. This technique is now able to more easily identify the outliers. Other techniques such as GMNP3 and GMNP5, which do have bounded influence, benefit less by the larger outlier magnitudes.

- The two high efficiency techniques (*M*-estimation and most B-robust) estimate worse than least squares. The fully iterated GM techniques (GMMZ and GMNP2) also do not estimate well, although they do improve over least squares
- Three techniques do a good job of estimating under these conditions; MM-estimation, GMNP3 and GMNP5

5.4.4.3 High Magnitude Outliers

Three scenarios are selected that represent a mix of problem sizes, outlier locations, outlier density, and leverage presence. The errors for the outliers are assigned large magnitudes ranging from 10 to 90. These large values greatly impact the least squares fit as evidenced by its AMSEE values of 7.6, 84.9 and 48.8.

- In terms of robust technique performance, nearly all of the methods estimate well. Only two techniques have a problem with the high leverage outlier, high outlier density problem.

M-estimation and most B-robust estimation have AMSEE values less than half that of least squares, but still unreasonable (38.7 and 25.0).

- Two techniques, MM-estimation and GMNP5, perform consistently better than the other robust techniques.

5.4.4.4 Outlier Location / Density

This three-factor, 24-run design generates some diverse robust technique performance results. The results in terms of AMSEE, robust technique ranking, percent over minimum AMSEE, and AMSIR values are presented in Table 5.12.

- As has occurred in all previous designs, no robust technique ranked first for all design runs. The best performing technique has an average rank of 3.0
- All of the robust techniques significantly improve accuracy of model estimation over least squares. Even LTS, which has an average rank of 9.3 out of 10 overall for the 24 datasets, has an AMSEE sum of 12.8, which is just 29% of the AMSEE sum for least squares (43.9)
- Several of the general comments made in the first experiment are also true in this case. *M*-estimation and most B-robust estimation perform poorly on datasets with high leverage outliers. Also, the multiple stage techniques perform better than the single stage methods

Table 5.12. Experiment 2.4 - Outlier Location / Density Performance Results

AMSEE												
DS	Description	LS	M	MB-R	LTS	S	MM	GMCH	GMMZ	GMNP2	GMNP3	GMNP5
1	2V, 12%, Int., No Lev	1.29	0.25	0.43	0.55	0.42	0.25	0.31	0.39	0.34	0.33	0.26
2	6V, 10%, Int, No Lev	1.63	0.21	0.32	0.61	0.49	0.20	0.28	0.30	0.28	0.27	0.22
3	10V, 10%, Int., No Lev	2.40	0.16	0.27	0.57	0.42	0.16	0.22	0.25	0.21	0.21	0.17
4	2V, 19%, Int., No Lev	2.54	0.31	0.47	0.49	0.37	0.33	0.39	0.43	0.48	0.43	0.30
5	6V, 20%, Int, No Lev	3.10	0.31	0.53	0.83	0.46	0.38	0.49	0.51	0.66	0.55	0.33
6	10V, 20%, Int., No Lev	4.85	0.17	0.40	0.63	0.34	0.17	0.32	0.37	0.44	0.39	0.17
7	2V, 12%, Int., Lev	0.63	0.11	0.16	0.31	0.20	0.12	0.17	0.20	0.20	0.17	0.17
8	6V, 10%, Int, Lev	0.41	0.08	0.11	0.46	0.29	0.08	0.15	0.13	0.13	0.12	0.15
9	10V, 10%, Int., Lev	0.53	0.04	0.07	0.36	0.22	0.06	0.12	0.09	0.09	0.09	0.11
10	2V, 19%, Int., Lev	0.81	0.11	0.16	0.31	0.15	0.13	0.14	0.22	0.25	0.16	0.15
11	6V, 20%, Int, Lev	1.04	0.08	0.14	0.43	0.21	0.09	0.18	0.23	0.34	0.19	0.17
12	10V, 20%, Int., Lev	0.87	0.05	0.08	0.28	0.14	0.06	0.13	0.10	0.16	0.10	0.09
13	2V, 12%, Ext, Lev	1.15	0.60	0.54	0.58	0.42	0.42	0.40	0.30	0.26	0.30	0.29
14	6V, 10%, Ext, Lev	1.50	0.78	0.84	0.68	0.44	0.26	0.35	0.24	0.21	0.25	0.24
15	10V, 10%, Ext, Lev	2.48	0.48	0.75	0.46	0.32	0.14	0.21	0.22	0.18	0.18	0.16
16	2V, 19%, Ext, Lev	2.03	1.29	0.95	0.65	0.49	0.86	0.54	0.32	0.34	0.39	0.38
17	6V, 20%, Ext, Lev	2.74	3.14	2.96	0.92	0.89	1.30	0.61	0.44	0.49	0.65	0.60
18	10V, 20%, Ext, Lev	4.81	6.05	5.14	0.68	0.52	0.28	0.36	0.34	0.39	0.44	0.31
19	2V, 12%, Int/Ext, Lev	0.83	0.23	0.32	0.42	0.28	0.16	0.24	0.26	0.19	0.20	0.21
20	6V, 10%, Int/Ext, Lev	0.98	0.53	0.46	0.64	0.44	0.19	0.30	0.24	0.20	0.22	0.24
21	10V, 10%, Int/Ext, Lev	1.29	0.13	0.22	0.40	0.27	0.08	0.17	0.15	0.13	0.13	0.14
22	2V, 19%, Int/Ext, Lev	1.78	1.33	0.96	0.52	0.27	0.45	0.40	0.38	0.39	0.26	0.20
23	6V, 20%, Int/Ext, Lev	1.93	1.39	1.07	0.63	0.39	0.50	0.41	0.33	0.38	0.32	0.26
24	10V, 20%, Int/Ext, Lev	2.73	1.11	1.02	0.46	0.23	0.18	0.24	0.26	0.30	0.23	0.15
	Sum	**44.36**	**18.95**	**18.39**	**12.86**	**8.67**	**6.84**	**7.15**	**6.74**	**7.04**	**6.60**	**5.49**

Robust Technique Ranking											
DS	Description	M	MB-R	LTS	S	MM	GMCH	GMMZ	GMNP2	GMNP3	GMNP5
1	2V, 12%, Int., No Lev	1	9	10	8	2	4	7	6	5	3
2	6V, 10%, Int, No Lev	2	8	10	9	1	5	7	6	4	3
3	10V, 10%, Int., No Lev	2	8	10	9	1	6	7	4	5	3
4	2V, 19%, Int., No Lev	2	8	10	4	3	5	7	9	6	1
5	6V, 20%, Int, No Lev	1	7	10	4	3	5	6	9	8	2
6	10V, 20%, Int., No Lev	2	8	10	5	1	4	6	9	7	3
7	2V, 12%, Int., Lev	1	3	10	9	2	5	7	8	6	4
8	6V, 10%, Int, Lev	2	3	10	9	1	7	6	5	4	8
9	10V, 10%, Int., Lev	1	3	10	9	2	8	5	6	4	7
10	2V, 19%, Int., Lev	1	7	10	5	2	3	8	9	6	4
11	6V, 20%, Int, Lev	1	3	10	7	2	5	8	9	6	4
12	10V, 20%, Int., Lev	1	3	10	8	2	7	6	9	5	4
13	2V, 12%, Ext, Lev	10	8	9	6	7	5	4	1	3	2
14	6V, 10%, Ext, Lev	9	10	8	7	5	6	3	1	4	2
15	10V, 10%, Ext, Lev	9	10	8	7	1	5	6	3	4	2
16	2V, 19%, Ext, Lev	10	9	7	5	8	6	1	2	4	3
17	6V, 20%, Ext, Lev	10	9	7	6	8	4	1	2	5	3
18	10V, 20%, Ext, Lev	10	9	8	7	1	4	3	5	6	2
19	2V, 12%, Int/Ext, Lev	5	9	10	8	1	6	7	2	3	4
20	6V, 10%, Int/Ext, Lev	9	8	10	7	1	6	5	2	3	4
21	10V, 10%, Int/Ext, Lev	3	8	10	9	1	7	6	2	4	5
22	2V, 19%, Int/Ext, Lev	10	9	8	3	7	6	4	5	2	1
23	6V, 20%, Int/Ext, Lev	10	9	8	5	7	6	3	4	2	1
24	10V, 20%, Int/Ext, Lev	10	9	8	4	2	5	6	7	3	1
	Sum of Ranks	**122**	**177**	**221**	**160**	**71**	**130**	**129**	**125**	**109**	**76**
	Std Dev of Ranks	**4.1**	**2.4**	**1.1**	**1.9**	**2.5**	**1.2**	**2.0**	**2.9**	**1.5**	**1.8**

Table 5.12. cont.

Percent Over Minimum AMSEE

DS	Description	M	MB-R	LTS	S	MM	GMCH	GMMZ	GMNP2	GMNP3	GMNP5
1	2V, 12%, Int., No Lev	0%	74%	118%	67%	2%	24%	56%	38%	30%	6%
2	6V, 10%, Int, No Lev	2%	58%	204%	143%	0%	39%	49%	41%	35%	10%
3	10V, 10%, Int., No Lev	2%	72%	266%	174%	0%	45%	64%	35%	36%	11%
4	2V, 19%, Int., No Lev	1%	53%	62%	20%	8%	28%	43%	58%	42%	0%
5	6V, 20%, Int, No Lev	0%	71%	168%	47%	24%	58%	64%	111%	77%	8%
6	10V, 20%, Int., No Lev	2%	139%	270%	100%	0%	91%	122%	162%	130%	3%
7	2V, 12%, Int., Lev	0%	47%	181%	81%	5%	55%	79%	81%	57%	49%
8	6V, 10%, Int, Lev	4%	35%	477%	267%	0%	93%	66%	66%	54%	96%
9	10V, 10%, Int., Lev	0%	65%	692%	387%	28%	173%	102%	110%	98%	150%
10	2V, 19%, Int., Lev	0%	53%	185%	40%	22%	31%	101%	130%	53%	40%
11	6V, 20%, Int, Lev	0%	77%	432%	165%	11%	125%	179%	316%	141%	115%
12	10V, 20%, Int., Lev	0%	56%	413%	158%	1%	143%	91%	190%	89%	71%
13	2V, 12%, Ext, Lev	126%	104%	118%	58%	59%	53%	15%	0%	13%	8%
14	6V, 10%, Ext, Lev	266%	293%	218%	107%	24%	66%	15%	0%	18%	12%
15	10V, 10%, Ext, Lev	236%	423%	223%	125%	0%	43%	52%	24%	24%	11%
16	2V, 19%, Ext, Lev	306%	199%	105%	54%	171%	70%	0%	6%	21%	18%
17	6V, 20%, Ext, Lev	607%	566%	106%	99%	192%	37%	0%	10%	46%	35%
18	10V, 20%, Ext, Lev	2076%	1750%	143%	88%	0%	28%	23%	40%	59%	13%
19	2V, 12%, Int/Ext, Lev	42%	97%	159%	71%	0%	46%	58%	18%	21%	27%
20	6V, 10%, Int/Ext, Lev	178%	143%	241%	133%	0%	58%	27%	4%	18%	26%
21	10V, 10%, Int/Ext, Lev	66%	186%	421%	242%	0%	123%	99%	65%	74%	80%
22	2V, 19%, Int/Ext, Lev	570%	386%	161%	37%	125%	104%	94%	97%	32%	0%
23	6V, 20%, Int/Ext, Lev	440%	314%	145%	50%	94%	59%	29%	48%	25%	0%
24	10V, 20%, Int/Ext, Lev	649%	584%	208%	58%	20%	59%	78%	100%	57%	0%
	Sum	**5571%**	**5845%**	**5717%**	**2772%**	**784%**	**1651%**	**1506%**	**1750%**	**1250%**	**789%**

AMSIR

DS	Description	M	MB-R	LTS	S	MM	GMCH	GMMZ	GMNP2	GMNP3	GMNP5
1	2V, 12%, Int., No Lev	0.19	0.34	0.42	0.32	0.20	0.24	0.30	0.27	0.25	0.21
2	6V, 10%, Int, No Lev	0.13	0.20	0.38	0.30	0.12	0.17	0.18	0.17	0.17	0.14
3	10V, 10%, Int., No Lev	0.07	0.11	0.24	0.18	0.06	0.09	0.11	0.09	0.09	0.07
4	2V, 19%, Int., No Lev	0.12	0.18	0.19	0.14	0.13	0.15	0.17	0.19	0.17	0.12
5	6V, 20%, Int, No Lev	0.10	0.17	0.27	0.15	0.12	0.16	0.16	0.21	0.18	0.11
6	10V, 20%, Int., No Lev	0.04	0.08	0.13	0.07	0.03	0.07	0.08	0.09	0.08	0.04
7	2V, 12%, Int., Lev	0.18	0.26	0.49	0.32	0.19	0.27	0.31	0.32	0.28	0.26
8	6V, 10%, Int, Lev	0.20	0.26	1.10	0.70	0.19	0.37	0.32	0.32	0.30	0.37
9	10V, 10%, Int., Lev	0.08	0.14	0.67	0.41	0.11	0.23	0.17	0.18	0.17	0.21
10	2V, 19%, Int., Lev	0.13	0.20	0.38	0.19	0.16	0.17	0.27	0.31	0.20	0.19
11	6V, 20%, Int, Lev	0.08	0.14	0.41	0.21	0.09	0.17	0.22	0.32	0.19	0.17
12	10V, 20%, Int., Lev	0.06	0.10	0.32	0.16	0.06	0.15	0.12	0.18	0.12	0.11
13	2V, 12%, Ext, Lev	0.52	0.47	0.50	0.36	0.37	0.35	0.26	0.23	0.26	0.25
14	6V, 10%, Ext, Lev	0.52	0.56	0.45	0.29	0.18	0.24	0.16	0.14	0.17	0.16
15	10V, 10%, Ext, Lev	0.19	0.30	0.19	0.13	0.06	0.08	0.09	0.07	0.07	0.06
16	2V, 19%, Ext, Lev	0.64	0.47	0.32	0.24	0.42	0.27	0.16	0.17	0.19	0.19
17	6V, 20%, Ext, Lev	1.15	1.08	0.34	0.32	0.47	0.22	0.16	0.18	0.24	0.22
18	10V, 20%, Ext, Lev	1.26	1.07	0.14	0.11	0.06	0.07	0.07	0.08	0.09	0.07
19	2V, 12%, Int/Ext, Lev	0.28	0.38	0.51	0.33	0.20	0.28	0.31	0.23	0.24	0.25
20	6V, 10%, Int/Ext, Lev	0.53	0.47	0.66	0.45	0.19	0.30	0.25	0.20	0.23	0.24
21	10V, 10%, Int/Ext, Lev	0.10	0.17	0.31	0.21	0.06	0.13	0.12	0.10	0.10	0.11
22	2V, 19%, Int/Ext, Lev	0.75	0.54	0.29	0.15	0.25	0.23	0.22	0.22	0.15	0.11
23	6V, 20%, Int/Ext, Lev	0.72	0.55	0.33	0.20	0.26	0.21	0.17	0.20	0.17	0.13
24	10V, 20%, Int/Ext, Lev	0.41	0.37	0.17	0.09	0.07	0.09	0.10	0.11	0.09	0.05
	Sum	**8.44**	**8.62**	**9.20**	**6.04**	**4.05**	**4.74**	**4.48**	**4.57**	**4.17**	**3.82**

5.4.5 Experiment 2 Individual Technique Performance

Summaries of technique performance are provided for each of the robust techniques tested. Strengths and weakness are discussed for the techniques' performance on the four sub-experiments. An effort is made to focus on additional information gained in this experiment over the first experiment. If differences in behavior occur, they will also be mentioned.

M-estimation

M-estimation is one of the three top performing techniques in terms of efficiency. The tuning constant used has the most impact on this behavior. The tuning constant value used (4.685) represents 95% efficiency relative to least squares at the normal error model.

M-estimation is also the preferred robust technique if the user is confident that the location of the outliers is in the interior **X**-space region. For the first 12 runs, which are all interior point outlier datasets, *M*-estimation has an average rank of 1.3. Unfortunately, if the user is wrong and outliers are located in leverage points, *M*-estimation is the worst of the robust techniques tested. This behavior also holds for the multiple point cloud datasets and the large magnitude outlier datasets. The AMSEE for two of the three cloud scenarios is significantly worse than least squares estimation. In the large outlier sub-experiment (2.3), the performance on the interior point outlier datasets is impressive, but is miserable on the exterior point outlier problem.

Most B-robust estimation

The weaknesses of this technique relative to other robust techniques outnumber its strengths. This method has reasonable efficiency, especially if good leverage points are present. It also does a fairly good job of fitting models with interior point outliers, although it does not rank in the top half of the robust techniques. Most B-robust estimation has trouble with exterior point outliers, particularly a large percentage of exterior point outliers. As a result, this technique does not properly estimate the cloud datasets and the exterior point large magnitude outlier dataset.

LTS

LTS is not an efficient technique relative to least squares when the subsample sizes are set for maximum breakdown performance. The subsample size for this and all preceding experiments is set so that LTS is a maximum or 50% breakdown estimator. This configuration results in AMSEE rankings of last on all six efficiency tests. Increases in efficiency can be obtained by sacrificing breakdown. It is important to note that, although this technique is unquestionably the worst performing of the robust methods tested, it consistently and significantly improves on least squares estimation.

S-estimation

S-estimation is better than LTS in efficiency, but significantly worse than the other robust methods. No significant performance differences occur as outliers are moved from the interior **X**-space to the exterior **X**-space. *S*-estimation is the only technique whose performance rank (relative to other techniques) is impacted by the size of the problem. Performance decreases with increasing number of model parameters. The average rank for all two-variable problems is 4.75, it is 7.00 for six-variable problems and 7.50 for 10-variable problems.

MM-estimation

MM-estimation performance in Experiment 2 is similar to its performance in the first experiment, which is outstanding. MM-estimation is one of the top two robust estimation methods tested. Very little additional information is gained in this second experiment. To reiterate the experiment one summary, MM-estimation is highly efficient, and works well against most types of outlier configurations. Its one softspot is datasets with a exterior point outliers. MM-estimation has more trouble with a large percentage (20%) of exterior point outliers. However, MM-estimation has the ability to accurately downweight high leverage outliers with large errors. This observation was noted in a previous section in the discussion on increasing the outlier magnitude of the multiple point cloud datasets. MM-estimation provides the second best fit to the large magnitude outlier dataset with exterior point outliers. This result indicates that MM-estimation's limited weakness is with moderate sized outliers (errors of 6-10 standard deviations) in high leverage points.

GM-estimation (New Proposal - 2)

This technique features an *S*-estimate initial with a fully iterated GM objective using a Huber ψ-function. This technique's combination of GM components results in a highly efficient estimator, as good as any robust technique tested. However, for the multiple point cloud datasets, the full GM iterations move the parameter estimates away from a descent initial estimate and towards a poor final estimate.

GM-estimation (New Proposal - 3)

This technique's performance is solid throughout. Although it is not one of the top two overall methods, it places a definite third. The description used in the first experiment results that it has no serious weaknesses can be restated here. It performs well in terms of efficiency, identifying and downweighting outlying clouds and high magnitude outliers, and is a good overall technique for fitting models to data with various outlier locations and densities.

GM-estimation (New Proposal - 5)

This technique is the multiple stage *S*-estimate initial approach followed by a one-step weighted least squares iteration of a GM objective with Tukey biweight ψ-function. The promising results in initial tests proved to be true indicators of its overall performance.

This method is one of the top two overall best performing robust methods. The only area this technique finished out of the top was the efficiency test. Regarding efficiency, GMNP5 falls in the second tier of robust methods. Its efficiency is significantly better than the high breakdown methods and comparable to most of the other GM techniques. In all three of the other sub-experiment areas, this method excels. It is one of the top techniques in detecting multiple point clouds and large magnitude outliers. GMNP5 is the overall best technique in estimating different outlier location / density and model dimension configurations.

GMNP5 was developed to improve on the performance of GMNP3 regarding interior point outliers. As a result, direct comparisons are made with GMNP3. The AMSEE numbers show that GMNP5 improves on or equals the performance of GMNP3 in every run. Improvements are concentrated on the interior point outliers without leverage points (DS1-DS8), and on the high outlier density, high leverage outlier datasets (DS16-18, DS22-24).

On the outlier location / density experiment, GMNP5 has the lowest AMSEE, second lowest overall rank, smallest rank standard deviation of the top techniques, smallest percent over minimum AMSEE, and the smallest overall AMSIR. It has an average rank of 3.2, and only three occurrences of a rank higher than four. No outlier location / density configurations identify a weakness in this method.

5.5 Checking the Adequacy of the Number of Replicates

The variability of the AMSEE values calculated for each design run is a function of the type of robust technique and the type of data being modeled. The most important information extracted from the experimental runs is the relative performance of the robust techniques. Understanding the variability of the AMSEE statistic is important, but so is determining whether the relative ranking of techniques will change by re-running the experiment. The variability of the AMSEE will be studied for two reasons. First, it is important to know which techniques have low estimates of AMSEE but a high standard error. These estimators should be viewed with caution because of the potential for occasional poor estimation. The other purpose in studying the AMSEE variability is to obtain a level of confidence in repeating the results if the experiment is run with different random number seeds.

To determine the level of confidence in the relative rank, three diagnostics of potential for change are obtained. The first metric consists of a correlation matrix for the robust techniques using a fifty replicate run. Also, two scenarios are repeated using a 50-replicate run and a 250-replicate run. The relative ranks for each scenario are compared. The third test consists of comparing two 16-run experiments in terms of the technique relative ranks for each run. Fifty replicates per run are used in this third test.

The correlation matrix is developed from several runs in experiment two and a representative one (DS2) is shown in Table 5.13. Many of the correlations are greater than 0.7, indicating a high level of correlation among techniques. The correlations of greatest importance are those between the top performing techniques, MM-estimation, GMNP3, and GMNP5. Those correlations (shaded) are all above 0.8.

Table 5.13. Correlation Matrix for an Experimental Run (Experiment 2, DS2)

	M	Most B-robust	LTS	S	MM	GMCH	GMMZ	GMNP2	GMNP3	GMNP5
M	1.00									
Most B-robust	0.73	1.00								
LTS	0.29	0.39	1.00							
S	0.68	0.60	0.52	1.00						
MM	0.99	0.77	0.32	0.70	1.00					
GMCH	0.74	0.81	0.71	0.67	0.78	1.00				
GMMZ	0.76	1.00	0.37	0.61	0.79	0.82	1.00			
GMNP2	0.74	0.84	0.27	0.41	0.78	0.80	0.86	1.00		
GMNP3	0.84	0.86	0.36	0.60	0.86	0.89	0.89	0.95	1.00	
GMNP5	0.95	0.70	0.37	0.82	0.95	0.76	0.73	0.64	0.81	1.00

To determine the impact that increasing the number of replicates has on relative rank, a pilot test is performed. Two datasets from Experiment 2 are compared separately using different replicate sizes. One run is performed using 50 replicates and the other using 250 replicates. The AMSEE ranks of the robust techniques are compared and studied for order switches. Table 5.14 displays the changes in ranks between replicate sizes.

Table 5.14. Relative Ranks of Robust Techniques Using Different Replication Sizes

DS4										
# of reps	M	RB Most	LTS	S	MM	GMCH	GMMZ	GMNP1	GMNP2	GMNP3
50	1	8	**10**	3	2	4	6	**5**	**9**	**7**
250	1	8	**9**	3	2	4	6	**7**	**10**	**5**
DS17										
50	10	9	8	**5**	7	**4**	1	6	3	2
250	10	9	8	**4**	7	**5**	1	6	3	2

The first run (DS4) shows two switches, one between ranks 5 and 7, and the other between 9 and 10. The second run shows only one switch. Only positions 4 and 5 are switched. The position changes that do occur are minor movements and take place among the lower ranking techniques which is of lessor consequence.

The final relative rank test consists of a complete experiment repeatability study. Experiment 1 was repeated after the decision to include GMNP3, so comparisons can be made on the original nine robust techniques. Table 5.15 shows the change in the relative ranks for each of the 16 runs. The rank differences are totaled to determine the overall impact of the repeated experiment on the rank sum measure.

The two most competitive techniques in this experiment, MM-estimation and GMNP2, have rank sum values that change by 3 and 1 respectively. Most of the cells have zeros in them indicating no change. The largest cell change is a 4 position switch which occurs only once. All of the other ranks change by 2 or fewer positions. By observing the AMSEE values in Table 5.15, there are several instances where differences in rank are decided by the third digit to the right of the decimal. Knowing that the standard deviation can be fairly high for the AMSEE values, this test result confirms that high correlations can have a positive impact in maintaining relative rank. A high level of confidence in the stability of the relative ranks is gained as result of this four-phase study.

Table 5.15. Experiment Repeatability Study - Differences in Run Ranks

Run	M	Most B	LTS	S	MM	GMCH	GMMZ	GMNP1	GMNP2
1	0	0	0	0	-2	0	0	1	1
2	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	1	-1	0	0
4	0	0	0	0	-1	1	0	0	0
5	0	-2	-2	0	4	0	0	0	0
6	0	-1	0	1	0	-1	1	0	0
7	-1	0	0	0	1	0	0	0	0
8	1	-1	-2	-1	-1	2	2	0	0
9	0	0	-1	0	0	0	1	0	0
10	1	-1	-1	0	0	1	0	0	0
11	0	0	-1	0	1	0	-1	0	1
12	-1	1	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	-1	1	0	0
15	0	0	0	0	0	0	1	0	-1
16	-1	1	-1	0	1	0	0	0	0
Total	**-1**	**-3**	**-8**	**0**	**3**	**3**	**4**	**1**	**1**

5.6 Conclusions

The two experiments performed on the robust regression methods considered in this study indicate that some robust methods do perform well across a variety of outlier and nonoutlier scenarios. The results also show that some robust methods are very specialized in the ability to perform well, while other methods (though significantly improve over least squares) perform consistently worse than the other robust methods.

The top two performing robust techniques consist of one that has been previously published (MM-estimation), and one that is introduced in this paper (GMNP5). The performance of both methods is impressive. Either method could be used with confidence on nearly any outlier or nonoutlier dataset configuration. If no outliers are present, MM-estimation is slightly preferred over GMNP5. However, in cases of high leverage outliers in small to moderate dimension models, the proposed GMNP5 technique is clearly superior to MM-estimation.

Chapter 6

A Robust, Robust Regression Technique

6.1 Introduction

The processes of exploratory data analysis and regression outlier detection are used to gain an improved understanding and characterization of data prior to model building. Among the many applications of these tools is their use as essential stepping stones in the model building process using regression analysis. Unfortunately, not everyone who uses regression analysis is properly trained or takes the time necessary to characterize data prior to fitting regression models. The increasing awareness of regression as a valuable tool and the dangerous combination of computer stored data and automated regression software have transformed regression modeling from the true science that it is, into a simple recipe of collect the data, fit the model, and use the equation.

The one estimation technique used overwhelmingly in regression analysis is least squares. The method of least squares is the uniform minimum variance unbiased estimator assuming the errors are normally distributed. Unfortunately, outliers are very common in data. Outliers can arise from simple computational or coding mistakes, by adding observations from a different population, or including odd response values due to machine failures or transient effects. Outliers can be present in a wide variety of densities, locations, magnitudes and groupings. Only one outlying observation can ruin least squares estimation, meaning that the parameter estimates do not provide useful information for the majority of the data.

Robust regression techniques have been developed to improve on least squares estimation and provide a compromise between fully including the outlier and throwing it away. The primary purpose of robust regression techniques is to fit models to the majority of the data. This general definition implies that these techniques should perform well on both messy data (with outliers), *and* on clean data (without outliers).

Not all robust techniques have the same effectiveness against both clean and messy data. Messy data can contain many different outlier configurations. Outlier configurations worthy of distinction are: a) outliers in the interior regressor or **X**-space versus outliers in the exterior or high leverage **X**-space, b) a small percentage of outliers versus a large percentage of outliers, c) outliers of moderate error magnitude versus large magnitude, and d) outliers dispersed as groups or clouds versus individual outliers. Robust techniques are evaluated and compared against each other both in their terms of their ability to properly fit the clean data, and how well they fit various outlier configurations.

The properties of efficiency, breakdown and bounded influence are used to define technique capability in a theoretical sense. Efficiency refers to the expected performance of a robust technique on clean data relative to the performance of least squares also on clean data. High efficiency techniques are desired. Breakdown is the percentage of outliers present in the data before the technique's parameter estimates are meaningless. For instance, least squares has a breakdown of $1/n$, indicating that only one outlier can render the estimates useless. High breakdown is preferred and some robust techniques have the maximum possible breakdown point of $n/2$ or 50%. Fifty percent breakdown means that up to half of the observations can be discrepant and the estimator will still provide useful information for the "good" portion of the data.. The third property, bounded influence, is designed to counter the tendency of least squares to allow points further out in the regressor space to exhibit greater influence on the parameter estimates. The purpose of bounding the influence is to reduce this exterior point influence, which can be especially important if these points are outliers.

Some of the more popular and well-respected robust methods include *M*-estimation, Least Trimmed Sums of Squares (LTS), *S*-estimation, MM-estimation, and Generalized *M*-estimation (GM-estimation). Of these techniques, only MM and GM-estimation have more than one desirable property. MM-estimation is a multi-stage technique that combines an initial *S*-estimate with a final *M*-estimate, resulting in an estimator that is both high efficiency and high breakdown. GM techniques are also multi-stage. GM-estimators have been proposed that are high efficiency, high breakdown, and bounded influence. These methods, proposed by Simpson et al. (1992) and Coakley and Hettmansperger (1993) combine certain high breakdown point initial estimates with other final estimation techniques, resulting in a three-property estimator. Many variations of GM-estimators are possible such that all three properties are maintained. A GM-estimate variant, quite different from the Simpson et al. and Coakley-Hettmansperger methods, is introduced in this paper. This estimator has high efficiency, bounded influence and demonstrates superior performance relative to the top performing robust methods, including these three-property GM-estimators. Monte Carlo simulation results show that this GM variant provides protection against all types of outlier scenarios and is also efficient. An estimator possessing these abilities is clearly a robust, robust method. A discussion of the technique components is provided along with explanations of the technique properties. Some comparisons of competing robust methods are provided using example datasets.

6.2 The Robust GM-estimator

The objective of GM-estimation, also known as bounded influence estimation, is to minimize a function of the residuals and leverage so that it is less sensitive to outliers than least squares. To be less sensitive to outliers, the objective function should increase less rapidly than the least squares sum of squared residuals function. To explain the GM approach, it makes sense to briefly discuss its predecessor, *M*-estimation. *M*-estimates are maximum likelihood estimators with the objectives of minimizing a function of the residuals. The objective can be expressed as

$$\min_{\beta} \sum_{i=1}^{n} \rho\left(\frac{e_i}{s}\right) = \min_{\beta} \sum_{i=1}^{n} \rho\left(\frac{y_i - \mathbf{x}_{\mathbf{i}}'\hat{\beta}}{s}\right) \tag{6.1}$$

Taking the partial derivatives of the objective with respect to β, and defining $\psi = \rho'$, the system of normal equations can be written as

$$\min_{\beta} \sum_{i=1}^{n} \psi\left(\frac{e_i}{s}\right) = \min_{\beta} \sum_{i=1}^{n} \psi\left(\frac{y_i - \mathbf{x}_{\mathbf{i}}'\hat{\beta}}{s}\right)\mathbf{x}_{\mathbf{i}} = \mathbf{0} \tag{6.2}$$

The *M*-estimator bounds the influence of the observation in the *y* direction (influence of the residual), but does not bound the influence of outliers in the regressor or X-space (influence of position). GM-estimators bound the influence of position in addition to the influence of residuals. GM-estimators weight the *M*-estimate system of equations

$$\sum_{i=1}^{n} \pi_i \psi\left(\frac{y_i - \mathbf{x}_i'\hat{\beta}}{s\pi_i}\right)\mathbf{x}_i = \mathbf{0} \tag{6.3}$$

The weights $\pi_i = \pi(\mathbf{x}_{ii})$, are a function of the measure of the distance $\mathbf{x}_i$ from the center of the multivariate regressor space cloud. The purpose of the π-weights is to enable reduced influence of the high leverage points. This particular type of objective function with the π-weights located inside and outside the argument of the ψ-function is called the Schweppe GM-estimator (Handshin et al. 1975). The other popular bounded influence function is the proposal of Mallows. The Mallows proposal has the π-weights located only outside the ψ-function argument. The Mallows objective tends to downweight high leverage observations regardless of the underlying residual, while the Schweppe objective attempts to downweight high leverage points only if the corresponding residual is large.

The computation of GM-estimates requires an initial estimate that produces a good starting point, followed by some type of iteration scheme that solves the nonlinear system of normal equations resulting in the final GM-estimate. The development of a GM technique involves a series of decisions regarding choices of initial estimators, estimators of scale, estimates of leverage, type of π-weight and ψ-function, type of nonlinear approach, and levels of convergence necessary. Table 6.1 outlines the components required to develop a GM-estimator and provides some general comments for each component.

Table 6.1. GM-Estimation Technique Characteristics

GM-Component	Comments
Bounded Influence Objective	The preferred type is that of Schweppe, who proposed an objective that theoretically downweights high leverage points only if the residual is large
Initial Estimate	The intent is to provide a good starting point. High breakdown estimators are typically used
Estimate of Scale	Several high breakdown choices are available including the MAD, the LMS estimate of scale, and the scale output of the initial estimate (from *S*-estimates). The scale estimate can be updated in final estimate iterations, but convergence is not necessarily assured
Estimate of Leverage	Different methods are available. A tradeoff exists between computational ease and the ability to handle clouds of multiple outliers
π-weights	Several different approaches are available corresponding to the type of leverage measure used. Some approaches require inlier/outlier cutoff values
ψ-function	*M*-estimate residual downweighting functions including Huber's *t*, Tukey's biweight and Ramsay's exponential
Tuning Constant (ψ-function)	Depends on the ψ-function and desired efficiency. Sometimes it is also a function of *n* and *p*
Convergence	Nonlinear convergence algorithms include, for example, Newton's method, and Iteratively Reweighted Least Squares (IRLS). Another consideration is the number of iterations so that the initial estimate breakdown property is preserved

6.3 GM-estimator Components

The previously published techniques by Simpson et al. and Coakley and Hettmansperger are methods with different GM-estimator components. The Coakley-Hettmansperger technique was introduced as an enhancement to the Simpson et al. method. For this reason, the Coakley-Hettmansperger method will be the method of these two used for comparison purposes. The GM-estimate variant proposed in this paper will be described in terms of its GM components (Table 6.2). The primary differences between this proposal and the two other GM methods are the types of initial estimate, measure of leverage, π-weights, and convergence technique used. In the following section, these components will be described in detail along with the reasons for selecting these approaches.

Table 6.2. Published and Proposed GM-estimate Robust Regression Techniques

Component	Technique		
	Simpson, Ruppert and Carroll (1992)	**Coakley and Hettmansperger (1993)**	**Proposal**
GM Objective	Mallows	Schweppe	Schweppe
Initial Estimate	LMS	LTS	*S*-estimate
Scale Estimate	MAD	$\hat{\sigma}_{LMS}$	$\hat{\sigma}_{S-est}$
Leverage Measure	Robust Distance (based on MVE)	Robust Distance (based on MVE)	Krasker-Welsch weights.
π-weight Function	min (1, b/RD^2)	min (1, b/RD^2)	median $\lvert z \rvert$ / $\lvert z \rvert$
ψ-function	Hampel	Huber	Tukey's Biweight
Tuning Constant	(a=1.5, b=3, c=8)	1.345	4.685
Convergence Approach	One-Step Newton-Raphson or Scoring	One-Step Newton-Raphson	One-Step RLS

6.3.1 Initial Estimate

The purpose of the obtaining an initial estimate in GM-estimation is to establish a good starting value so that subsequent estimation stages can bound the influence and increase efficiency. The ideal type of preliminary estimate is one with high breakdown, so that the final solution can possibly maintain this additional property. Various high breakdown estimators have been used for this purpose, namely the LMS and LTS estimators. As shown in Table 6.2, these estimators are used for the Simpson et al. and Coakley-Hettmansperger approaches. Another possible high breakdown estimator is *S*-estimation. Each of these high breakdown estimators can be configured at 50% breakdown. The efficiencies of these estimators configured at such a breakdown vary substantially. LMS and LTS have 0% and 7.1% asymptotic efficiencies and *S*-estimators have 28.7% efficiency. Although the *S*-estimator efficiency is still somewhat low, it is significantly higher than LMS or LTS. The higher efficiency of *S*-estimators is the primary reason for its selection as the proposed initial estimate.

The objective functions for LMS and LTS are

$$\text{LMS:} \quad \min_{\beta} med\, e_i^2 \tag{6.4}$$

$$\text{LTS:} \quad \min_{\beta} \sum_{i=1}^{h} (e^2)_{i:n} \tag{6.5}$$

where $(e^2)_{1:n} \leq (e^2)_{2:n} \leq \ldots \leq (e^2)_{n:n}$ are the ordered squared residuals and h is the number of residuals included in the calculation. This approach is similar to least squares except the largest α squared residuals are not used (trimmed sum) in the summation, allowing the fit to avoid the outliers. For *S*-estimation the objective is

$$S\text{-estimation:} \quad \min_{\beta} \; s\big(e_1(\beta), \cdots, e_n(\beta)\big) \tag{6.6}$$

The dispersion function $s\big(e_1(\beta), \cdots, e_n(\beta)\big)$ is found as the solution to

$$\frac{1}{n-p} \sum_{i=1}^{n} \rho\left(\frac{y_i - \mathbf{x}_i'\hat{\beta}}{s}\right) = K \tag{6.7}$$

The constant K may be defined as $E_{\Phi}[\rho]$, where Φ stands for the standard normal distribution.

For LMS and LTS, the objective functions are nondifferentiable. For LMS, LTS, and *S*-estimates, the objective functions may have many local minima. So, algorithms based on differentiating the objective function or based solely on local improvement (e.g. Newton's method) are not appropriate for optimizing these estimators. Therefore, numerical procedures for the computation of all three high breakdown estimates require global optimization routines. Global techniques often require some type of random search approach. Random subsampling procedures, which are invariant to reparameterization, are used for each of these high breakdown methods. Random subsampling is computationally intensive, especially as the size of the problem increases. Fortunately, Ruppert (1992) has developed an algorithm for *S*-estimation that requires less evaluations of the objective function resulting in faster convergence than LMS or LTS estimation.

S-estimation is also asymptotically normal (see Rousseeuw and Yohai 1984) meaning that hypothesis tests can be performed on the parameter estimates and asymptotic efficiencies can be computed for various values of tuning constants. Efficiency can be increased at the expense of decreases in breakdown point. Some situations may warrant this type of tradeoff and subsequent selection of appropriate values for K and c. Table 6.3 lists some alternatives to the high breakdown point constants.

Table 6.3. Breakdown Point (ε*) and Asymptotic Efficiency (e) of *S*-estimators Using Tukey's Biweight Function and Various Combinations of Constants c and K

ε^* (%)	e (%)	c	K
50	28.7	1.548	0.1995
45	37.0	1.756	0.2312
40	46.2	1.988	0.2634
35	56.0	2.251	0.2957
30	66.1	2.560	0.3278
25	75.9	2.917	0.3593
20	84.7	3.420	0.3899
15	91.7	4.096	0.4194
10	96.6	5.182	0.4475

Source: Table 19 of Rousseeuw, P. J., and Leroy, A. M. (1987), *Robust Regression and Outlier Detection*, Wiley, N. Y.

6.3.2 Measures of Leverage

The bounded influence property of GM-estimators depends on the π-weights being adequate measures of regressor space leverage. As it is stated, this problem only deals with the observations' **X**-space location. The basic approach of techniques designed to measure leverage is to determine the center of **X**-space mass and then to use some means for measuring the distance each point lies from the center of mass. Like the regression estimators discussed in this paper, we would like the measures of leverage to be robust. In other words, it is desired that points outlying in the regressor space not influence the location of the center of mass and the resulting distance measures. Traditional regression measures of leverage are the hat diagonals, which are main diagonal elements of $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. It has been shown by several authors (e.g. Rousseeuw and van Zomeren 1990), that this measure is not very robust. High leverage points can cause fairly significant changes in the measure of the location of the centroid and thus in the resulting distances. Improvements to this approach have been offered in the design and development of *M*-estimates of covariance, the Minimum Volume Ellipsoid (MVE) estimator and the Minimum Covariance Determinant (MCD) estimator. The *M*-estimates of covariance (see Hampel et al. 1986) are robust, but have breakdown limited to $1/p$. The MVE and MCD estimators (Rousseeuw 1983, 1984) have breakdowns as high as 50% but suffer from computational problems and have some outlier identification vulnerabilities. *M*-estimates of covariance and the two high breakdown methods will be discussed in more detail to clearly point out the benefits and pitfalls of each method.

M-estimates of covariance were first suggested by Hampel (1973) but the basic paper on these estimators is attributed to Maronna (1976). Maronna addressed the problems of existence, uniqueness, asymptotic distribution and breakdown point. To compute *M*-estimates of covariance, first consider the classical measure of covariance

$$\hat{\mathbf{C}} = \left(\sum_i \mathbf{x}_i \mathbf{x}_i' \right)^{-1} \tag{6.8}$$

Let $\hat{\mathbf{A}}$ be a *p* x *p* transformation matrix that makes the data spherically symmetric, such that

$$\sum_{i=1}^{n} z_i z_i^T = \mathbf{I}_p \tag{6.9}$$

where $z_i = \hat{\mathbf{A}}\mathbf{x}_i$. and $\mathbf{I}_p$ is a *p* x *p* identity matrix. So $\hat{\mathbf{C}} = (\hat{\mathbf{A}}' \ \hat{\mathbf{A}})^{-1}$ is not resistant to outliers, but the influence can be bounded by introducing weights $u(|z_i|)$

$$\sum_{i=1}^{n} u(|z_i|) z_i z_i' \ = \mathbf{I}_p \tag{6.10}$$

and $u(|z|)$ is some decreasing function. The choice of $u(|z|)$ is different depending on the type of estimator. For the Krasker-Welsch (KW) approach, $u(|z|) = \xi(c/|z|)$ and

$$\xi(c/s) = (c/s)^2 + (1-(c/s)^2)(2\Phi(c/s)-1) - 2(c/s)\phi(c/s) \tag{6.11}$$

where c is a given constant, Φ represents the standard normal cumulative distribution and ϕ is the density of Φ. Hampel et al. (1986) show that a necessary condition for the existence of the transformation matrix **A**, is that the tuning constant $c > \sqrt{p}$. We can now find the *M*-estimates of covariance and location defined by the system of equations from (6.10) and

$$\sum_{i=1}^{n} w(|z_i|) z_i = 0 \tag{6.12}$$

where $z_i = \hat{\mathbf{A}}(x_i - \hat{t})$, $\hat{t}$ is an estimate of location and w is a suitable weight function. For the Krasker-Welsch approach, $w(|z|) = 1 / |z|$. So, using the Krasker-Welsch method and understanding that the distances for the *M*-estimates of covariance are $|z|$, the π-weights are the w-function, which are just the inverses of the distances.

Huber (1977, 1981) and Stahel (1981) show that the breakdown point of affine *M*-estimators of covariances is at most $1/p$, which can be a problem in large dimension data. Fortunately, in some instances these estimators work well above their breakdown point. Consider a 10-variable scenario consisting of points arranged in a fractional factorial setting with axial points. The fractional factorial is a 2^{10-4} design (64 points) with 16 axial points used to represent high leverage positions. The percentage of high leverage points is then 16/80 or 20%. The theoretical breakdown of the *M*-estimate of covariance is $1/p \cong 9\%$. Computing *M*-estimate (KW-type) covariance distances correctly identifies the axial points as high leverage (Table 6.4).

Table 6.4. Krasker-Welsch Distances for the 10-Variable, 20% Outlier Dataset

Case(s)	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	*M*-estimate Distance
1-64	±1	±1	±1	±1	±1	±1	±1	±1	±1	±1	17.0
65	10	0	0	0	0	0	0	0	0	0	48.9
66	0	12	0	0	0	0	0	0	0	0	58.4
67	0	0	14	0	0	0	0	0	0	0	68.2
68	0	0	0	16	0	0	0	0	0	0	77.8
69	0	0	0	0	10	0	0	0	0	0	48.9
70	0	0	0	0	0	12	0	0	0	0	58.4
71	0	0	0	0	0	0	14	0	0	0	74.7
72	0	0	0	0	0	0	0	16	0	0	85.3
73	0	0	0	0	0	0	0	0	10	0	53.5
74	0	0	0	0	0	0	0	0	0	12	64.1
75	-14	0	0	0	0	0	0	0	0	0	68.2
76	0	-16	0	0	0	0	0	0	0	0	77.8
77	0	0	-10	0	0	0	0	0	0	0	48.9
78	0	0	0	-12	0	0	0	0	0	0	58.4
79	0	0	0	0	-14	0	0	0	0	0	68.2
80	0	0	0	0	0	-16	0	0	0	0	77.8

The robust distances based on the MVE and MCD estimators, have a breakdown of 50% if the subsample sizes are chosen appropriately. The robust distances are computed using the Mahalanobis equation, substituting robust measures of location and covariance for the classical measures. The Mahalanobis distance (*MD*) is computed using

$$MD_i = \sqrt{(\mathbf{x}_i - T(\mathbf{X}))C(\mathbf{X})^{-1}(\mathbf{x}_i - T(\mathbf{X}))'} \qquad (6.13)$$

A robust alternative to *MD*, proposed by Rousseeuw (1983, 1984), is to replace *T*(**X**) and *C*(**X**) with robust measures of multivariate location and dispersion. He suggested using the MVE and MCD to generate the robust measures. The MVE is a generalization of the LMS regression technique for multivariate location/dispersion. The estimate of location is the center of the minimum volume ellipsoid covering half of the observations. The covariance matrix is computed using the same ellipsoid and a correction factor to obtain consistency at multinormal distributions. The idea for MCD is to use the concept of LTS to develop these measures. The approach yields

$$T(\mathbf{X}) = \text{Mean of the } h \text{ points of } \mathbf{X} \text{ for which the determinant of the covariance matrix is minimal} \tag{6.14}$$

Thus, if we define an empirical mean and covariance matrix based on a subset of points $\mathbf{H}_n$ of size h of $\{1,\ldots,n\}$, to be $T(\mathbf{H}_n)$ and $C(\mathbf{H}_n)$,

$$T(\mathbf{H}_n) = \frac{1}{h}\sum_{i=1}^{h}\mathbf{x}_i\,, \qquad C(\mathbf{H}_n) = \frac{1}{h}\sum_{i=1}^{h}(\mathbf{x}_i - T(\mathbf{H}_n))(\mathbf{x}_i - T(\mathbf{H}_n))' \tag{6.15}$$

Then, the subset $\hat{\mathbf{H}}_n$ of $\{1, \ldots, n\}$ for which the determinant of $T(\mathbf{H}_n)$, $|T(\mathbf{H}_n)|$, attains its minimum value over all subsets $\mathbf{H}_n$. is the matrix used for computation of the MCD.

The MCD corresponds to finding the h points for which the classical tolerance ellipsoid has minimum volume and then locating its center. The MCD also yields a robust covariance matrix which is found by multiplying the classical covariance matrix of the selected h points by a constant to obtain consistency in the case of a multivariate normal distribution.

6.3.3 Testing the Leverage Measures

Some cloud datasets are developed to test the robust distances using the MVE estimator. Each dataset consists of a two-variable design of 12 interior points (3 replicates of a 2^2 factorial design), and 4 high leverage points in a cloud located some place away from the 12-point cube. The total design space consists of 16 points, 25% of which are **X**-space outliers. The purpose of the study is to identify vulnerabilities in the leverage estimation methods. The goal is to find locations for the 4-point cloud that are significantly far away from the cube, but not easily detected by one or more of the location/dispersion techniques. The MVE robust distances estimator used in the tests was a 50% breakdown estimator. However, using datasets with only 25% high leverage points, the technique failed to detect some outlying cloud locations. Remembering that the 50% breakdown MVE estimator is designed to identify the minimum volume ellipsoid covering just over half of the data, it is possible to locate the high leverage points far away from the cube but within the MVE. Placing the cloud in-line with points in the cube results in an minimum volume ellipse that is elongated and narrow, covering the high leverage points. The diagram below illustrates this scenario.

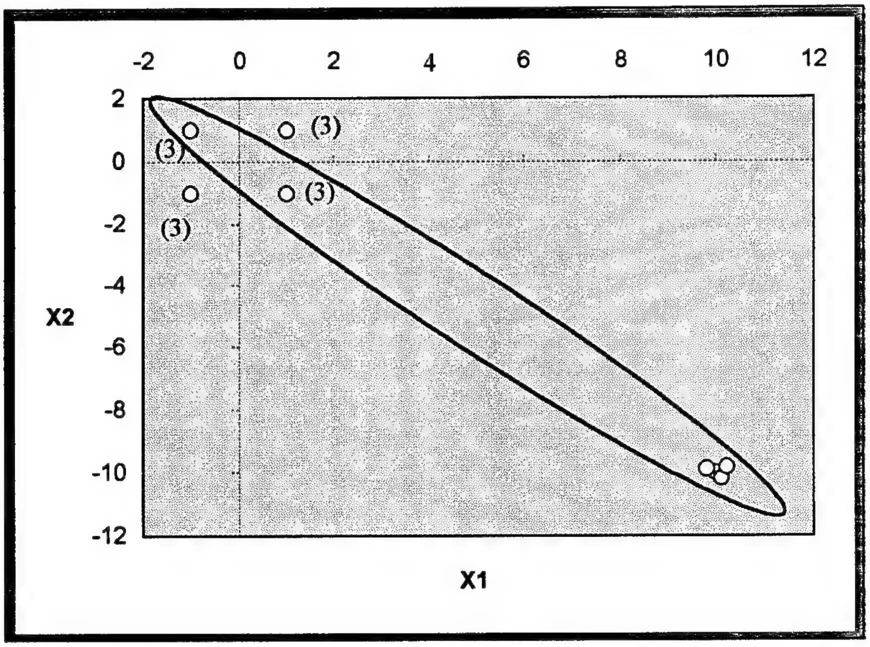


Figure 6.1. Minimum Volume Ellipsoid Covering Outlying Point Cloud

Not only does the MVE technique "mask", or fail to correctly identify the discrepant cloud, but it also "swamps" the off-diagonal cube elements, meaning that inliers are incorrectly specified as outliers. In this example, these off-diagonal cube points have robust squared distances that are nearly fifteen times the cutoff value suggested by Rousseeuw and van Zomeren (1991).

Another cloud location using this 2-variable design reveals a different dynamic of the MVE concept. Consider placing the cloud (not all points identical) along one of the axes and steadily increasing its distance from the cube. Figure 6.2 shows the different subsample points selected for the MVE as the cloud moves away from the cube. Due to the alignment of the cloud relative to subset points in the cube, the MVE technique does not identify the cloud as outliers until those points are moved a considerable distance from the cube points, indicating perhaps that the power of this technique may not be high for this type of scenario.

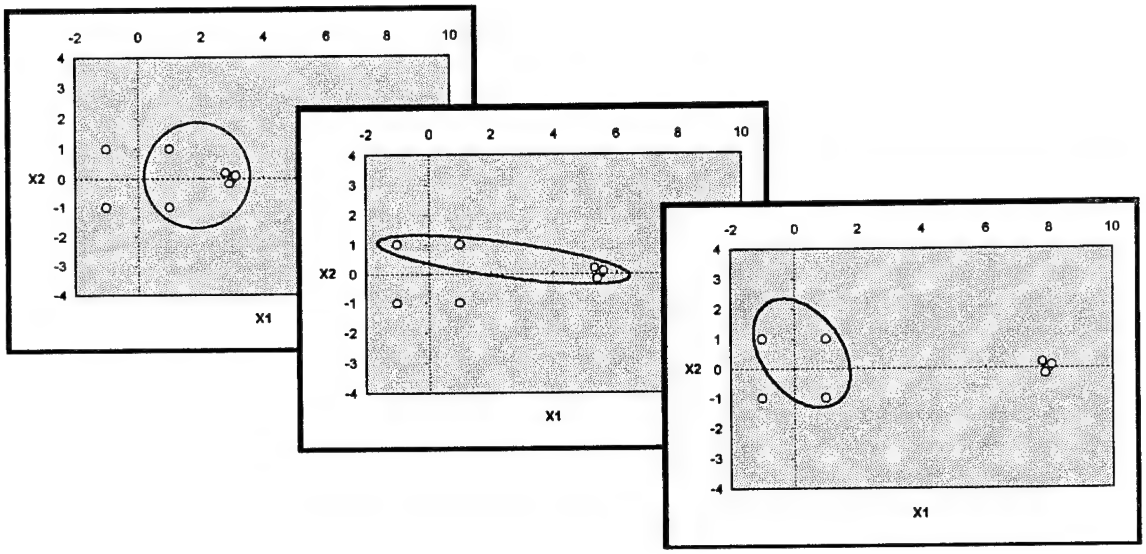


Figure 6.2. Minimum Volume Ellipsoid Dynamics as Cloud Moves Along X1 Axis

Other techniques, such as the Krasker Welsch distances, also have trouble correctly identifying the outlying cloud as being positioned significantly far away from the rest of the points.

Because the cloud forces the centroid away from the majority of points toward the cloud, the cloud points have only slightly larger distances than the cube points furthest from the centroid. However, the Krasker Welsch weights represent improvements over MVE distances because huge distances are not assigned to actual inliers (off-diagonal cube points) as they are in the MVE approach. Clearly, all of the diagnostics have weaknesses which indicates the difficult nature of the multiple point cloud problem.

The MVE robust distances approach requires a global optimization routine using random sampling of point subsets. The approach consists of drawing random subsamples of $p + 1$ observations. In order to perform a complete or exhaustive search of all possible subsamples, $\mathrm{M} = \binom{n}{p+1}$ combinations are required. For example, a dataset with 6 variables and 40 observations requires inspection of over 18 million subsamples. Obviously, selecting a random subset of subsamples is preferred in such a case. Unfortunately, Cook and Hawkins (1990) point out that this approach can be inefficient because random subsets require even more draws to ensure that the minimum volume ellipse is found. Random search typically requires M log M random sets be drawn to find the minimum. As a result, sometimes even large subsamples will not result in an estimate close to the true minimum, thus large variability in these approximations is often experienced. One more caution is reported by Cook and Hawkins (1990), Simonoff (1991), and Hettmansperger and Sheather (1991). They all note that the MVE approach has a tendency to identify outliers ("false positives") in clean data. Simonoff performed Monte Carlo simulations on data using MVE robust distances. High leverage points were identified as observations with distances exceeding the χ^2 (α=0.025, p-1) cutoff. Simonoff found that, on the average, the robust distances technique leads to 5 out of 20 cases being misidentified as outliers in clean data.

The outlook for the MCD estimator does look better than the MVE. The one similarity to MVE is its performance on the cloud datasets. The MCD identifies inliers as outliers, and outliers as inliers. Fortunately though, the computational aspects of the MCD have improved. Hawkins (1994) has developed algorithms for the MCD that have resulted in stable approximations and computation times significantly faster than the MVE approximations. The MCD is also asymptotically normal (Butler et al. 1993) at normal distributions, increasing the potential for developing reasonable hypothesis tests and prediction intervals.

Although *M*-estimates of covariance do not theoretically have a high breakdown for large problems, the technique does generate accurate, stable estimates for a large percentage of scenarios, even for some configurations with outlier densities beyond their theoretical breakdown. The author regards the shortcomings of this approach as less severe than those of the high breakdown methods. For this reason, the *M*-estimates of covariance method is used to measure the leverage in the proposed GM-estimator.

6.3.4 The π-weights

The π-weights corresponding to the Krasker-Welsch method are computed as the inverse of the distances for each observation. This approach is simple and intuitive. Unfortunately, implementation of this method in a GM-estimator reveal that the π-weights are in many cases small for interior **X**-space points. The resulting argument of the ψ-function is large resulting in excessive downweighting. Both Krasker and Welsch (1982) and Walker (1984) consider the amount of overall downweighting a significant factor in the effectiveness of a GM-technique. The preferred estimator is the one that downweights only the discrepant observations. Any unnecessary downweighting is not only undesirable, but results in lower efficiency relative to least squares under normal errors.

A modification is made to the π-weights in an attempt to reduce the overall downweighting and to increase the efficiency. Instead of using the inverse of the distances, the numerator of the π-weight expression becomes the median distance observation. So the π-weights become

$$\pi_i = \underset{j}{\mathrm{med}}|z_j|/z_i \tag{6.16}$$

which reduces the magnitude of the π-weights significantly while maintaining the same relative impact. The high leverage points result in π-weight<1, increasing the argument of the ψ-function. Monte Carlo results (Chapter 4) show this to be an effective improvement over the Krasker-Welsch weights.

6.3.5 The ψ-function

Although there are many specific proposals for the ψ-function, they can all be grouped into one of two classes: monotone and redescending. The monotone methods (e.g. Huber's) continue to downweight observations of increasing argument magnitude, while not assigning zero weight to an observation unless the argument is zero. These methods are stable in terms of convergence properties. The redescending functions (e.g. Tukey's biweight) will assign zero weight to ψ-function arguments (residual and π-weight combinations) beyond a certain value. Thus, if it is suspected that large outliers or outliers in high leverage points are present, the redescending functions will be more robust. Convergence problems can occur with redescending functions, so care should be taken that a good initial estimate is obtained. Holland and Welsch (1977) and Birch (1980) assert that, for redescending functions, only local convergence can be assured.

An important finding in Monte Carlo studies (Chapter 4) is that, for some outlier situations, even small nonzero weights can result in undesirable parameter estimates. The ψ-function used with the proposed GM-estimator is the redescending Tukey's biweight. The initial estimate provides a good starting point and only one reweighted least squares step is used, so convergence is not an issue.

6.3.6 Convergence Technique

Solving the system of nonlinear equations formed by taking the partial derivatives of the Schweppe objective function requires some sort of iterative technique. Several iterative approaches have been developed including iteratively reweighted least squares (IRLS), Newton's method, and a technique proposed by Huber (1973) referred to as the H-algorithm. In general, each of these techniques performs well. IRLS is the most popular because any least squares program can be used to solve the equations. Newton's method is regarded as the most efficient, requiring the least number of iterations to converge. One consideration in a multi-stage robust estimator is the characteristics of the estimator after the final stage is implemented. Fully iterated convergence may not maintain some of the desirable properties of estimates from earlier stages.

Those interested in maintaining more of the properties of the initial estimate came up with the idea of limiting the number of steps of the iterative algorithm. Rousseeuw (1984) proposed this idea for IRLS, suggesting a one-step reweighted least squares (RLS) approach be used after an initial LMS estimate. The purpose of the technique is to increase the efficiency while maintaining the high breakdown characteristics of the initial estimate. One proposal tested in the simulation experiment is a robust-ridge technique that uses a one-step RLS method.

The solution technique for the proposed GM technique is a one-step RLS method. The RLS is easy to implement. The square root of the weights are multiplied by the responses and regressors and an ordinary least squares program can solve for the parameter estimates. The weights have interpretive and diagnostic value because the observations with less than full weight have reduced influence or are omitted from the model (zero weight). Monte Carlo simulation (Chapter 4) also indicates that the one-step method takes substantial steps toward improving the final estimate. A final benefit of this approach is that it can be easily adapted to incorporate ridge regression if the regressor variables are highly collinear.

6.3.7 Robust Regression Inference

Several authors have studied the problem of inference regarding the GM-estimate parameter coefficients. Maronna and Yohai (1981) show that fully iterated GM-estimates are asymptotically normal with normal covariance matrix. Simpson et al. (1992) caution that one-step reweighted least squares results in a final estimate that inherits the asymptotics of the initial estimate. In the case of the proposed GM methods, the initial *S*-estimates are asymptotically normal causing the final estimate to be asymptotically normal.

Ronchetti (1982), Markatou and Hettmansperger (1990), Birch and Agard (1993), Markatou and He (1994), and Heritier and Ronchetti (1994) have all developed or compared GM-estimate inference tests that can be used to study the significance of regression and the significance of the individual parameter estimates. Useful tests include a class of τ tests introduced by Ronchetti (1982) for GM-estimates, the aligned generalized M test of Markatou and Hettmansperger (1990), and other Wald-type tests (see Hampel et al. 1986, p. 363). Some of these tests, including the class of τ tests, have been implemented and are available in ROBETH (see Marazzi 1993).

6.4 Proposed Method Performance

The performance of the proposed GM-estimator will be studied using simulation results and an example. The technique will be compared to highly regarded robust methods relative to efficiency, breakdown and bounded influence. One of the performance statistics used to evaluate techniques is the mean square error of estimation. This statistic is appropriate for simulation studies where the true model coefficients are known. The mean square error of estimation is then found by

$$\text{MSEE} = (\hat{\beta}_R - \beta)'(\hat{\beta}_R - \beta) \tag{6.17}$$

where $\hat{\beta}_R$ is a vector of robust technique parameter estimates and β is the vector of true model coefficients. Simulation experiments are often replicated to reduce the variability in the performance statistic, so the *average* mean square error of estimation (AMSEE) is often computed.

6.4.1 Efficiency

Robust technique efficiencies are determined relative to least squares and can be found in an asymptotic theoretical sense for many methods, given their tuning constant settings. Another means of evaluating efficiencies is to compare performances of techniques relative to least squares and each other using data from the target or normal distribution. An experiment is performed using a design matrix of regressors and errors generated from random normal distribution variates. The number of parameters and presence of leverage points are varied so that changes in estimation behavior can be studied. Larger dimension problems may cause problems for some robust techniques. The presence of "good" leverage points should improve most techniques' accuracy of estimation due to the heavy influence of these leverage points. The GM-estimates may not improve as much due to their influence bounding behavior. Maronna, Yohai and Zamar (1993) observe that the efficiency of GM-estimators depends on the spread of the regressor variables, and may be arbitrarily low for heavy-tailed distribution of **x**'s.

The datasets used for the efficiency demonstration are two, six, and ten variable factorial or fractional factorial designs. The ratio of sample points to variables is about 8:1. Each dataset has 80% of the points on the corners of the factorial cuboidal region and 20% on axial points. When leverage points are required, the axial points are moved a significant distance away from the cube centriod, so that the datasets have 20% high leverage points. The responses are generated from known and fixed parameter coefficients and a random normal variate error value. The model signal-to-noise ratio is about 100:1 for each sized design. Fifty replicates of each combination of model dimension and leverage presence are run to obtain average MSEE values for each technique.

Several comparative robust methods are used, including known high and low efficiency techniques. *M*-estimation uses the Tukey biweight ψ-function with tuning constant of 4.685, representing the 95% efficiency level. LTS and *S*-estimation subsample sizes and tuning constants are set for 50% breakdown and 7.12% and 28.7% efficiencies, respectively. MM-estimation uses the final stage tuning constant of 4.687 which gives 95% efficiency. The Coakley-Hettmansperger method employs the Huber ψ-function with c=1.345 for 95% efficiency. Two variations of the proposed GM technique are tested. One proposal (GMP-T) uses the Tukey biweight ψ-function (c=4.685) and the other (GMP-H) uses the monotone Huber ψ-function (c=1.345). Both tuning constants are set for 95% efficiency. Table 6.5 lists the AMSEE values for each design run.

Table 6.5. Robust Technique Efficiency Performance (AMSEE)

Dataset	LS	M	LTS	S	MM	GMCH	GMP-T	GMP-H
2V, No Lev	0.20	0.22	0.74	0.55	0.21	0.28	0.25	0.24
6V, No Lev	0.19	0.21	0.80	0.69	0.21	0.31	0.29	0.25
10V, No Lev	0.13	0.16	0.60	0.53	0.15	0.21	0.20	0.18
2V, Lev	0.09	0.11	0.53	0.38	0.12	0.15	0.22	0.15
6V, Lev	0.07	0.08	0.47	0.38	0.10	0.16	0.19	0.13
10V, Lev	0.04	0.05	0.34	0.28	0.06	0.12	0.13	0.09
Sum	**0.71**	**0.82**	**3.48**	**2.81**	**0.85**	**1.23**	**1.29**	**1.04**

6.4.2 Breakdown

The breakdown of various robust methods is normally measured for a certain subsample size or tuning constant value and is fixed for any dataset characteristic. However, for GM techniques, breakdown is a function of the number of model parameters. To expose GM techniques to high breakdown situations, a large number of outliers needs to be present in large dimension problems. Monte Carlo simulation is again used to test robust method performance. Two of the designs used in the efficiency test are adapted for this exercise. The two complaints regarding the original GM methods are that their breakdown is only $1/p$ for detecting outliers, and $1/p$ for detecting high leverage points. Enhancements to the original methods have resulted in the techniques discussed, such as the Coakley-Hettmansperger method which is high breakdown. However, the proposed GM method is restricted to the breakdown of the leverage estimation technique, $1/p$. Therefore, to fully challenge the proposed methods, the two designs used in this test have their percentage of outliers *and* percentage of leverage points > $1/p$. The six-variable dataset (p=7, including the intercept) has 20% leverage points and 20% high leverage outliers, which is greater than $1/p$=14%. The 10-variable dataset (p=11) has the same 20% leverage and 20% high leverage outliers, which is also greater than $1/p$=9%. The same performance statistic, AMSEE, based on fifty replicate runs is computed to compare estimation accuracies. Table 6.6 shows the results for the same methods described in the efficiency test.

Table 6.6. Robust Technique Breakdown Performance (AMSEE)

Dataset	LS	M	LTS	S	MM	GMCH	GMP-T	GMP-H
6V, 20% Lev, 20% Outlier	2.74	3.14	0.92	0.89	1.30	0.61	0.60	0.65
10V, 20% Lev, 20% Outlier	4.81	6.05	0.68	0.52	0.28	0.36	0.31	0.44
Sum	**7.55**	**9.19**	**1.70**	**1.41**	**1.58**	**0.97**	**0.91**	**1.09**

The results indicate that the proposed GM methods perform well overall and do not experience breakdown in estimation accuracy, even though the percentage of high leverage points and outliers is beyond their theoretical breakdown point. In fact, the Tukey version of the proposed method (GMP-T) performs the best overall. The Coakley-Hettmansperger method performs next best, followed by the Huber version of the new proposal.

6.4.3 Bounded Influence

An example will be used to demonstrate to ability of the robust methods to fit the majority of the data. Because the outliers in the example are also high leverage points, the condition requires techniques to be able to bound their influence. The same robust methods used previously will be fit to the data.

An Example

The science of cost estimation often involves developing models for data containing outliers. The business of estimating costs for government satellites is no exception. In this situation, a limited number of data points is available, restricting the number of parameters that can be used to generate a model with adequate degrees of freedom for error. Experts in this field determined that the best predictor of first unit cost is overall satellite weight. For a particular class of satellites, 19 observations, each representing a satellite, are collected. The actual data has been modified slightly to challenge the bounded influence aspects of the estimators. A plot of the data (Figure 6.3) reveals that four observations are not in-line with the other 15 observations. The specific data values are provided in Table 6.7.

Table 6.7. Satellite Cost Data

Observation	Cost ($K)	Weight
1	2449	90.6
2	2248	87.8
3	3545	38.6
4	794	28.6
5	1619	28.9
6	2079	23.3
7	918	21.1
8	1231	17.5
9	3641	27.6
10	4314	39.2
11	2628	34.9
12	3989	46.6
13	2308	80.9
14	376	14.6
15	5428	48.1
16	2786	38.1
17	2497	73.2
18	5551	40.8
19	5208	44.6

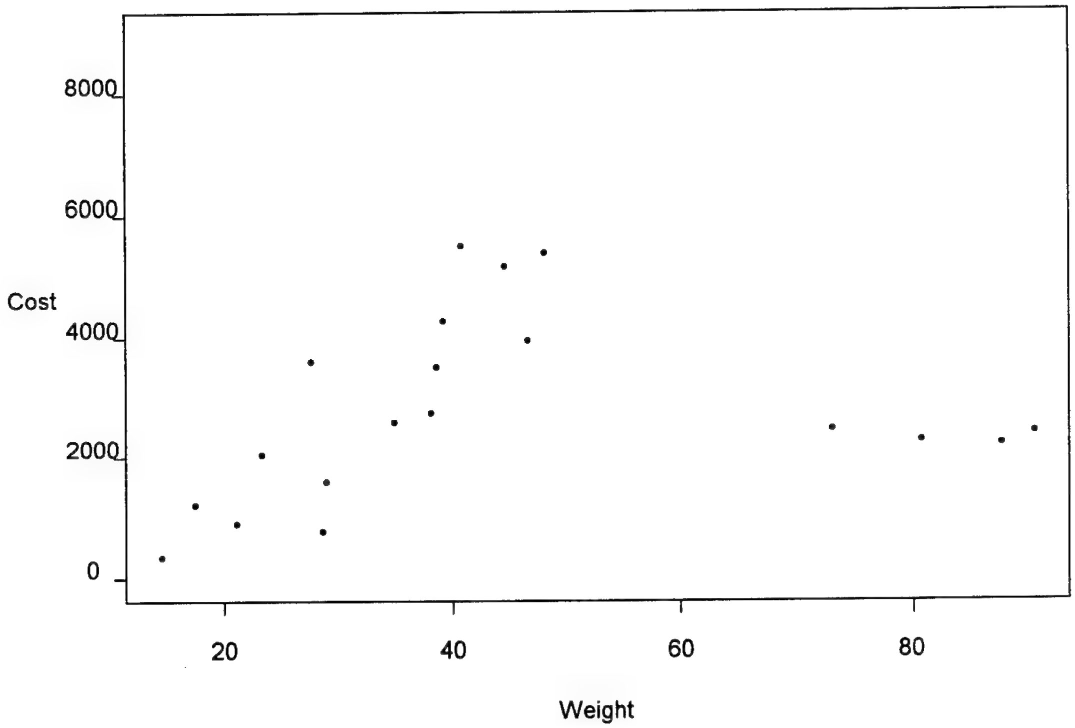


Figure 6.3. Modified Satellite Cost Estimation Data

It is anticipated that least squares and some of the robust methods without bounded influence will have trouble fitting the majority of the data because the outliers are high leverage points, outside the range in **X**-space from the other points. The high leverage points will tend to pull the regression line in their direction.

The resulting least squares and robust fits are displayed together in Figure 6.4. Only the proposed GM methods avoid being pulled totally away from the bulk of the data. Although each of the GM techniques use the Schweppe objective function which is intended to bound the influence of high leverage outliers, not all the GM techniques succeeded in that endeavor in this example. The

Coakley-Hettmansperger method performed no better than least squares, probably because the initial LTS estimate is so far from the true fit. Only one Newton convergence step is used to obtain the final Coakley-Hettmansperger estimates. Even though the robust distances (using the MVE) correctly identifies the four high leverage points, one iteration of the bounded influence procedure is not sufficient. The difference between the GMP-T and GMP-H techniques lies in the power of the redescending ψ-function in GMP-T to drive the outliers to zero weight. The resulting GMP-T fit to the remaining 15 observations is preferred over the GMP-H line.

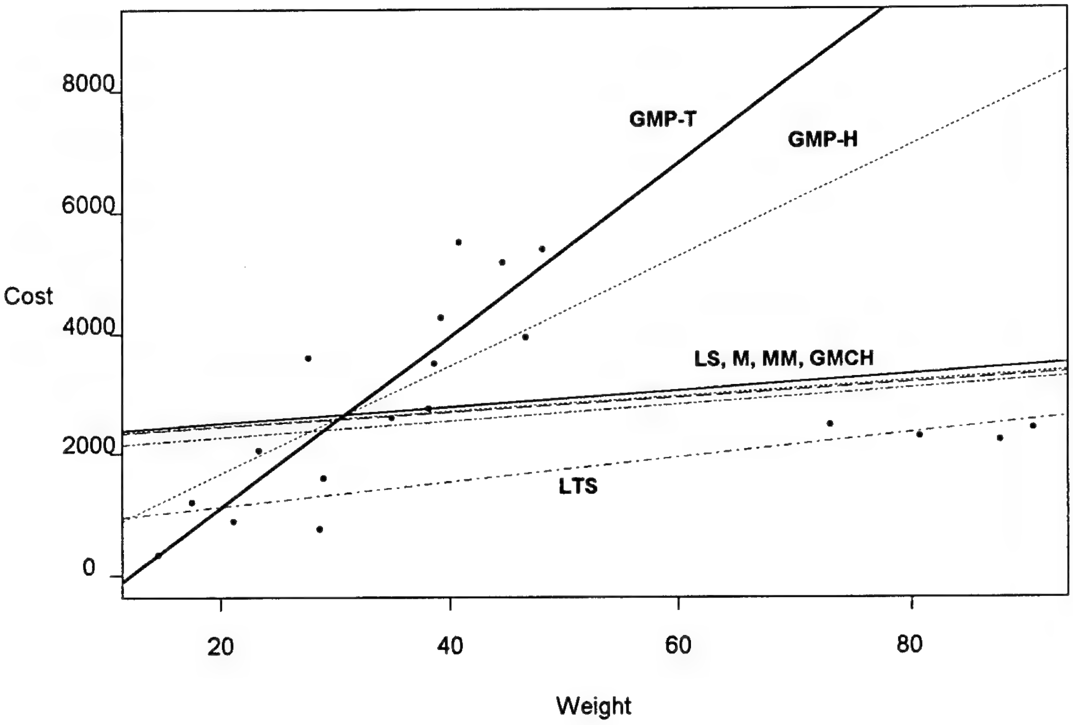


Figure 6.4. Satellite Cost Data with Robust Regression Estimates

LTS also did a poor job of fitting the data. The line drawn using the LTS parameter estimates makes sense by recalling the LTS objective. The idea of LTS is to minimize the squared residuals for a subset of h points. For the 50% breakdown estimator, $h \cong 10$. If you look closely at the LTS line, there are 10 points in the vicinity of the line. Due to the spread of the 15 nonoutliers, there probably is not a line running in that direction that has a smaller sum of squared residuals for 10 observations. This type of situation really highlights one of the weaknesses of the high breakdown methods.

The fits for the other robust methods, *M* and MM-estimation fail because their objective functions cannot bound the influence of the high leverage outliers. The problem with MM-estimation is clearly with the final stage *M*-estimate. Both MM and the two GMP methods use *S*-estimation in the initial stage, so the differences in the lines indicates the difference between an unbounded and bounded final stage.

The author used S-PLUS in conjunction with the libraries of ROBETH (Marazzi 1993) to develop the proposed GM-estimates as well as all of the other robust methods. The code for the proposed technique GMP-T is provided in Appendix B. The S-PLUS commands are available from the author upon request.

6.5 Conclusions

The proposed GM techniques have been described in detail and their performance is demonstrated in three areas. GMP-T and GMP-H have high efficiency, bounded influence, and demonstrate high breakdown. In terms of efficiency, a simple way to improve estimation accuracy relative to least squares under normal error conditions is to modify the tuning constant used in the *S*-estimate initial stage. Huber (1993) states that it is difficult to find rationale for a breakdown point higher than 25%, if it would result in a serious efficiency loss at the model, or in excessive computation times. A 25% breakdown *S*-estimator would increase its asymptotic relative efficiency from 28.7% (at 50% breakdown) to 75.9%. Monte Carlo studies (Chapter 5) have shown that MM-estimation and GMP-T perform the best overall against a comprehensive set of outlier conditions. MM-estimation has trouble with high leverage outliers in small to moderate dimension data. The satellite data is a good demonstration of MM-estimate's weakness. GMP-T does not have any clear weaknesses and performs near the top of the robust methods in a variety of scenarios. If it is important that all observations receive nonzero weight, the GMP-H method can be used as a viable alternative. Either method can also be computed easily using available software. The final weights can be reported and used to better understand the data.

The study of robust methods has been and will continue to be a very important topic for those in search of methods that fit the bulk of the data regardless of the presence of outliers. With data collection and regression analysis automation a reality, a robust fitting alternative must be available in situations where automated fitting is performed. The additional benefit of a good robust technique for use as a diagnostic tool is also a possibility.

Many robust techniques have been developed, proposed, and studied intensively in terms of their properties of efficiency, breakdown and bounded influence. Several authors have reported simulation studies, but few have been comprehensive assessments of technique performance against a wide variety of outlier scenarios. In addition, new techniques are being developed at rapid rates, so relevancy in comparative studies requires that they be current. This proposed GM method with either the Tukey or Huber ψ-function, has been demonstrated to perform well in comparative situations against the better robust methods. The correlation between expected performance (theoretical) and demonstrated performance (empirical) is high for these methods, but findings sometimes conflict. However, the true value of a robust method for the user lies in its ability to perform well empirically.

Chapter 7

A Robust-Ridge Regression Technique for the Combined Outlier-Multicollinearity Problem

7.1 Introduction

Two commonly occurring assumption violations that continually threaten effective least squares regression estimation are empirical realities leading to dependencies among the regressor variables and a lack of normally distributed error terms. Although least squares estimation is fairly robust to slight departures from these assumptions, it is possible that only a single large outlier can render least squares estimation meaningless. For example, a single observation that departs significantly from the fit of the other observations can pull a least squares model in its direction and result in an overall fit to none of the data points. Likewise, regression models with as few as two linearly dependent parameters can alter a least squares estimate so that it is not only off by an order of magnitude, but it can actually switch sign. Most real regression datasets have a measurable degree of both maladies; nonnormal errors often caused by outliers and linearly dependent regressors referred to as multicollinearity. Outliers can occur for reasons of computation or coding errors, mistaken pooling of observations from separate populations, or equipment testing failure. The group of methods designed to be less sensitive than least squares to outliers and which offer a compromise to valid outlier deletion, are robust regression techniques.

To accommodate the linear dependency among regressors and provide improved estimation, biased estimation techniques such as ridge regression, sacrifice small amounts of bias for large reductions in estimator variance. Many datasets suited for regression estimation contain outliers, and many of the appropriate regressors are correlated such that the least squares estimate can be significantly improved by the use of robust and/or biased estimation methods.

These weaknesses of least squares and their associated remedies have garnered the attention and energies of many research efforts, primarily in a separate treatment of the issues. Far fewer efforts have focused on solutions to the frequently occurring simultaneous problem. Holland (1973) was the first to propose a solution for the combined problem which is a ridge method applied to a robust estimator. Askin and Montgomery (1980) proposed an augmented robust method allowing for a family of biased methods to be combined with the popular *M*-estimation robust technique. Unfortunately *M*-estimation has trouble downweighting high leverage points, or points outlying in the regressor space. So, Walker (1987) proposed augmented bounded-influence estimators that followed the Askin-Montgomery approach but replaced *M*-estimation with generalized *M*-estimators, which appropriately downweight high leverage points.

The purpose of this paper is to introduce a new augmented robust method that improves on the robustness of previous methods and to demonstrate its performance. A description of these previous methods will be provided in Section 7.2 prior to a discussion in Section 7.3 of the components of the proposed approach. Section 7.4 contains a presentation of a Monte Carlo experiment developed and performed to test previous methods with the proposed method. An example is introduced in Section 7.5 that is a modification of the Hawkins, Bradu, and Kass (1984) dataset. The top performing methods are evaluated using this data and results are discussed. Section 7.6 offers some conclusions and future directions for research in this area of study.

7.2 Evolution of Robust-Ridge Regression Methods

Consider the linear regression model of the form

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{7.1}$$

where **y** is a n x 1 vector of response observations, **X** is an n x p matrix of the levels of the regressor variables, $\boldsymbol{\beta}$ is a p x 1 vector of model coefficient estimates, and $\boldsymbol{\varepsilon}$ is an n x 1 vector of errors. The errors are assumed to have $E(\varepsilon) = 0$ and $V(\varepsilon) = \sigma^2$, and are assumed to be uncorrelated. The classic least squares estimates for the model coefficients is given by the solution to

$$\boldsymbol{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \tag{7.2}$$

When the model errors are normally distributed the method of least squares estimation is attractive in the sense that the estimate of $\boldsymbol{\beta}$ has desirable statistical properties. If we assume that the errors are normally distributed, we can show that the least squares estimates are also the maximum likelihood estimates and that these estimates are the uniformly minimum variance unbiased estimators. Under conditions of nonnormal distributions, particularly heavy-tailed error distributions, least squares no longer has these desirable properties. In fact, it can be shown that the maximum likelihood estimator for the heavy-tailed Laplace distribution is a robust technique called the least absolute value (LAV) estimator. In general, robust methods seek to minimize some function of the residuals that is less sensitive to outliers. Insensitivity is obtained by replacing the least squares objective (sum of the squared residuals) with a function of the residuals that is less gradual, perhaps of the form

$$\min_{\beta} \sum_{i=1}^{n} \rho(y_i - \mathbf{x}_i^{'} \hat{\beta}) \tag{7.3}$$

The function ρ is a function of the residuals that is often related to the likelihood function for an appropriate error distribution. The class of estimates using this approach is *M*-estimation. A number of ρ functions have been proposed and although several perform well in general, no variation is universally the best. The objective function is normally minimized by using an iterative technique to solve the resulting nonlinear system of normal equations of the form

$$\sum_{i=1}^{n} \psi(e_i / s)\mathbf{x}_i = \mathbf{0}$$

where ψ is the derivative of ρ and $\mathbf{x}_i$ is the row vector of explanatory variables of the i^{th} case. An estimate of scale s is needed to standardize the residuals because the solution to the normal equations is not equivariant with respect to a magnification of the y-axis. The most common approach for solving this system is iteratively reweighted least squares (IRLS), resulting in an estimator of the form

$$\hat{\beta} = (\mathbf{X'WX})^{-1}\mathbf{X'Wy}$$

where $\mathbf{W} = diag(w_1, w_2, \ldots, w_i)$ and $w_i = \psi(e_i / s) / (e_i / s)$. This approach is widely used because any ordinary least squares package can be used to perform the estimation.

The other concern with least squares is the potential for multicollinearity among the regressors. Multicollinearity, or a near linear dependence among regressors, can result in an **X′X** matrix that is ill-conditioned, meaning that it is difficult to accurately invert. The computed least squares estimates are often too large in magnitude and the variance of $\hat{\beta}$ may be very large. The remedy proposed by Hoerl and Kennard (1970a, 1970b), called ridge regression, is to induce a slight bias which will shrink the parameter estimates and significantly reduce the variance of $\hat{\beta}$. Ridge estimation can be performed by slightly modifying the least squares normal equations to be

$$(\mathbf{X}'\mathbf{X} + k\mathbf{I})\hat{\beta}_R = \mathbf{X}'\mathbf{y} \tag{7.4}$$

so,

$$\hat{\beta}_R = (\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}'\mathbf{y} \tag{7.5}$$

This estimate tends to shrink the least squares estimate in absolute value with respect to the contours of $\mathbf{X}'\mathbf{X}$. So, $\hat{\beta}$ can be viewed as the solution to

$$\underset{\beta}{\text{Minimize}}\ (\beta - \hat{\beta})\mathbf{X}'\mathbf{X}(\beta - \hat{\beta}) \tag{7.6}$$

subject to $\hat{\beta}'\hat{\beta} \le d^2$ where d^2 depends on k.

In terms of the combined outlier-collinearity problem, some significant contributions have been made by Holland (1973), Pariente and Welsch (1977), Askin and Montgomery (1980), Pfaffenberger and Dielman (1985), and Walker (1987). Holland was the first to propose a robust-biased estimation technique by suggesting the use of a weighted ridge estimation, with a robust choice of weights. He suggested first performing robust estimation to obtain the weights followed by ridge regression on the weighted data. Pariente and Welsch decided to constrain least absolute value regression and solve the system of equations using linear programming. Askin and Montgomery offered an approach for combining a number of biased estimation techniques with robust estimation. They implement Marquardt's (1970) suggestion to augment the regressor matrix with a diagonal matrix of the biasing parameter(s) for the biased estimation portion. They then perform iteratively reweighted least squares on the augmented matrix using robust weights. Specifically, Askin and Montgomery propose robust-ridge estimators that are the solution to the problem

$$\min_{\beta} \sum_{i=1}^{n} \rho(y_i - \mathbf{x}_i' \hat{\beta}) \tag{7.7}$$

$$\text{subject to } \hat{\beta}' \hat{\beta} \leq d^2$$

where the objective function is the classic *M*-estimator described previously. The solution to this combined problem, obtained by IRLS, requires that the weights on augmented observations be fixed at 1.0 so that the integrity of the augmented matrix is maintained. The augmented matrix serves an important function in stabilizing the otherwise ill-conditioned **X′X** matrix. The resulting estimator becomes

$$\hat{\beta} = (\mathbf{X}'\mathbf{W}\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}'\mathbf{W}\mathbf{y} \tag{7.8}$$

where $\mathbf{W} = diag(w_1, w_2, \ldots, w_i)$ and $w_i = \psi(e_i / s) / (e_i / s)$ as in *M*-estimation above. Using this approach of first finding the biasing parameter and then performing robust estimation, the weighting matrix is a function of the shrinkage parameter *k*. This distinction is important based on the findings of Walker and Birch (1988), who show that a relationship exists between influence and collinearity. In particular, they find that the influence of each observation depends largely on the value of *k*. Askin and Montgomery not only demonstrated their method on actual data, but also performed an extensive sensitivity study using Monte Carlo simulation to determine which combination of robust and ridge techniques performed the best. Several *M*-estimation ψ-functions were compared as well as different biased estimation techniques. Although no combination was universally superior, they found that the combination of the *M*-estimate using the Hampel ψ-function with ridge estimation had the best overall performance. The Askin-Montgomery approach represents a significant step forward in robust-ridge estimation, but their approach using robust *M*-estimation does not have bounded influence, resulting in poor estimation performance when points are outlying in the **X**-space.

A natural extension of augmented *M*-estimators are augmented bounded-influence estimators (Walker 1987). Bounded influence estimators, also known as generalized M or GM-estimators, were introduced to bound the influence of points outlying in the **X**-space by means of a weight function. Points for which the $\mathbf{x}_i$ is far away from the bulk of the $\mathbf{x}_i$ in the data are called high leverage points. Schweppe (Handshin et al. 1975) proposed an objective function with associated normal equations of the form

$$\sum_{i=1}^{n} \psi(\frac{y_i - \mathbf{x}_i^{'}\hat{\beta}}{\pi_i s})\pi_i \mathbf{x}_i = \mathbf{0} \tag{7.9}$$

where the π_i are a function of $\mathbf{x}_i$, and are included to provide an indication of **X**-space leverage. Several approaches are available for computing these π-weights. These approaches are primarily characterized by the method used to measure the leverage; some methods are more robust than others. The bounded influence, biased estimation approach proposed by Walker is then the solution to

$$\min_{\beta} \sum_{i=1}^{n} \rho(\frac{y_i - \mathbf{x}_i^{'}\hat{\beta}}{\pi_i s})\pi_i \tag{7.10}$$
$$\text{subject to } \hat{\beta}'\hat{\beta} \leq d^2$$

where the objective function is now the bounded-influence estimation approach. The estimator is found using (7.8) above, but in this case the weights are found by applying the bounded-influence approach where $w_i = \psi\ (e_i / \pi_i\ s) / (e_i / \pi_i\ s)$. The weights in this case are not fixed, but are also functions of the shrinkage parameter *k*.

Walker suggested using the *DFFITS* measure for the π-weights, which includes the h_{ii} (hat diagonals) as the measure of leverage. Unfortunately, the h_{ii} are not very robust when multiple high leverage points are present. Rousseeuw and van Zomeren (1990) point out that the h_{ii} are influenced by high leverage points and the center of mass used to measure the leverage distance is drawn in their direction. The resulting h_{ii} values are not necessarily large for the high leverage points, so the diagnostic does not correctly indicate the leverage. This failure to correctly identify outliers as outliers is known as the masking effect. Several examples are provided by Rousseeuw and van Zomeren, demonstrating the masking tendency of the h_{ii} metric.

The Walker method also suggests using least squares as the initial estimate of the robust estimation scheme. A good initial estimate in bounded influence regression is important because the iteration schemes will have a tendency to converge to a local minimum sometimes not necessarily close to the global optimum. Robust methods that are insensitive to multiple outliers, called high breakdown estimators, are the preferred choice for bounded influence initial estimates.

Some of the robust aspects of Walker's robust-ridge estimator warrant investigation for possible improvements. The components of the combined estimator are important determinants of its estimation ability. An alternative robust-ridge estimator is introduced and its components will be discussed in the next section. A discussion of the proper sequence for dealing with the combined problem is also included, as it also can have substantial impacts on estimation success.

7.3 The Proposed Robust-Ridge Estimator

The general method proposed by Askin and Montgomery and followed by Walker for performing robust-ridge estimation has been demonstrated to work well in general, especially for the multicollinearity aspect of the problem. Possible vulnerabilities of these methods lie in their treatment of the outlier condition. For this reason, the differences between the proposed method and previous methods lie primarily in the robust estimation components. The components of the proposed method will be described in this section. Also, modifications to the sequence of estimation are suggested and will be discussed.

The proposed estimator is similar to the Walker proposal in that it is an augmented bounded influence approach. The normal equations and equations for the parameter estimates are the same as those given in (7.8 and 7.9). The differences are found primarily in the choices for the initial estimate and the π-weights for the robust estimator. The sequence used by both Askin and Montgomery, and Walker to develop the combined estimation consists of the following steps:

1. Obtain initial model estimates for determination of the ridge biasing parameter. Parameter estimates ($\hat{\beta}$) and mean square error ($\hat{\sigma}^2$) are required.
2. Estimate the ridge regression biasing parameter. The technique preferred by many is analysis of the ridge trace. Alternatives include the one-step proposal of Hoerl, Kennard, and Baldwin (1975), and the iterative technique proposed by Hoerl and Kennard (1976). The analytical methods have the advantage that they may be automated.
3. Add the augmented **X** matrix, using the estimated biasing parameter, to the original **X** matrix.
4. Perform augmented robust estimation using IRLS. The initial weights are normally set to one and converge to their final values through iteration. The weights for the augmented observations are kept at one.

This sequence is desired if the problems of collinearity need to be addressed prior to robust estimation. However, in order to determine the biasing parameter, an initial estimate is needed. Many have suggested using a robust estimate of $\hat{\beta}$ and $\hat{\sigma}^2$. Askin and Montgomery use least squares. But Walker (1984) mentions that for his applications he used his proposed GM-estimator with the *DFFITS* ψ-function argument for these initial estimates. This means that his robust-ridge estimator actually involves two GM-estimates, which may not necessarily be the best approach.

Two proposals are offered in this paper that both start with robust initial estimates to generate the biasing parameters for ridge estimation. The first proposal then uses the initial estimate robust weights in augmented weighted least squares to obtain the final estimate. The second proposal involves re-estimating the robust weights in augmented IRLS. Both of these proposals eliminate a number of computational steps from Walker's method and satisfy the steps outlined above.

7.3.1 Initial Estimate

The purpose of the initial estimate in GM-estimation is to obtain a good starting value that can be improved by bounding the influence and increasing efficiency. The ideal type of preliminary estimate then is one with high breakdown. Breakdown is the percentage of outliers present in the data before the technique's parameter estimates are unreliable or meaningless. For instance, least squares has a breakdown of $1/n$, indicating that only one outlier can render the estimates useless. High breakdown is preferred and some robust techniques have the maximum possible breakdown point of $n/2$ or 50%, so that the final solution can maintain this property. Various high breakdown estimators have been used as starting points for bounded influence estimation, namely the LMS and LTS estimators. These estimators are used for the proposed approaches of Simpson et al. (1992) and Coakley and Hettmansperger (1993). Another possible high breakdown estimator is *S*-estimation. Each of these estimators can be configured as 50% breakdown techniques. The efficiencies of these estimators configured at maximal breakdown vary considerably. LMS and LTS have 0% and 7.12% asymptotic efficiencies and *S*-estimators have 28.7% efficiency. Although the *S*-estimator efficiency is relatively low, it is significantly higher than LMS or LTS. It is selected primarily for this reason.

The objective function for *S*-estimation is

$$\min_{\beta} \; s\big(e_1(\beta),\cdots,e_n(\beta)\big) \tag{7.11}$$

where the dispersion function $s\big(e_1(\beta),\cdots,e_n(\beta)\big)$ is found as the solution to

$$\frac{1}{n-p}\sum_{i=1}^{n}\rho\left(\frac{y_i-\mathbf{x}_i'\beta}{s}\right)=K \tag{7.12}$$

The constant *K* may be defined as $E_{\Phi}[\rho]$, where Φ represents the standard normal distribution.

For *S*-estimates, the objective functions may have many local minima, making algorithms based solely on local improvement (e.g. Newton's method) not appropriate for optimizing these estimators. Therefore, numerical procedures for the computation of these high breakdown estimates require global optimization routines. Global techniques often require some type of random search approach. Random subsampling procedures, which are invariant to reparameterization, are used for each of these high breakdown methods. Random subsampling is computationally intensive, especially as the size of the problem increases. Fortunately, Ruppert (1992) has developed an algorithm for *S*-estimation that requires less evaluations of the objective function resulting in faster convergence than LMS or LTS estimation.

S-estimation is also asymptotically normal (see Rousseeuw and Yohai 1984) meaning that hypothesis tests can be performed on the parameter estimates and asymptotic efficiencies can be computed for various values of tuning constants. Efficiency can be increased at the expense of decreases in breakdown point. Some situations may warrant this type of tradeoff and subsequent selection of appropriate values for *K* and *c*.

7.3.2 Measures of Leverage

The bounded influence property of GM-estimators depends on the π-weights being adequate measures of regressor space leverage. As it is stated, this problem only deals with the observations' **X**-space location. The basic approach of techniques designed to measure leverage is to determine the center of **X**-space mass and then to use some means for measuring the distance each point lies from the center of mass. Like the regression estimators discussed in this paper, we would like the measures of leverage to be robust. In other words, it is desired that points outlying in the regressor space not influence the location of the center of mass and the resulting distance measures. Traditional regression measures of leverage are the hat diagonals, which are main diagonal elements of $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. It has been shown by several authors (e.g. Rousseeuw and van Zomeren 1990), that this measure is not very robust. High leverage points can cause fairly significant changes in the measure of the location of the centroid and thus in the resulting distances. Improvements to this approach have been offered in the design and development of *M*-estimates of covariance, the Minimum Volume Ellipsoid (MVE) estimator and the Minimum Covariance Determinant (MCD) estimator. The *M*-estimates of covariance (see Hampel et al. 1986) are robust, but have breakdown limited to $1/p$. The MVE and MCD estimators (Rousseeuw 1983, 1984) have breakdowns as high as 50% but suffer from computational problems and have some outlier identification vulnerabilities. The MVE and MCD estimators are computed using a global optimization search routine that considers subsamples on p+1 points. For moderate to large sized problems, random subsampling is required. Unfortunately, the random subsampling does not guarantee finding the true minimum. Several algorithms have been developed to calculate the approximate and exact solutions. In general the exact solutions take too long and the approximate solutions have considerable variability. An exact solution for a 6 regressor, 40 observation dataset requires over 18 million evaluations, which required over eight hours of processing on a 486DX2-66 PC. The author also performed a study of the performance of different leverage measures against datasets with clusters of outliers. On several occasions the high breakdown methods not only masked the points in the outlier cluster, but these methods also identified some of the inliers as outliers, a behavior known as swamping. Other authors, including Cook and Hawkins (1990), Simonoff (1991), and Hettmansperger and Sheather (1991), note that the MVE approach has a swamping tendency. Simonoff performed Monte Carlo simulations on data using MVE robust distances. High leverage points were identified as observations with distances exceeding a χ^2 (α=0.025, p-1) cutoff. Simonoff found that, on the average, the robust distances technique leads to 5 out of 20 cases being identified in clean data. For these reasons, the two high breakdown methods were not used for measuring leverage distances. The selected technique, *M*-estimates of covariance, will be discussed in more detail.

M-estimates of covariance were first suggested by Hampel (1973) but the basic paper on these estimators is attributed to Maronna (1976). Maronna addressed the problems of existence, uniqueness, asymptotic distribution and breakdown point. To compute *M*-estimates of covariance, first consider the classical measure of covariance

$$\hat{\mathbf{C}} = \left(\sum_i \mathbf{x}_i \mathbf{x}_i' \right)^{-1}$$

Let $\hat{\mathbf{A}}$ be a p x p transformation matrix that makes the data spherically symmetric, such that

$$\sum_{i=1}^{n} z_i z_i^T = \mathbf{I}_p$$

where $z_i = \hat{\mathbf{A}}\mathbf{x}_i$. and $\mathbf{I}_p$ is a p x p identity matrix. So $\hat{\mathbf{C}} = (\hat{\mathbf{A}}'\ \hat{\mathbf{A}})^{-1}$ is not resistant to outliers, but the influence can be bounded by introducing weights $u(|z_i|)$

$$\sum_{i=1}^{n} u(|z_i|) z_i z_i' = \mathbf{I}_p \tag{7.13}$$

and $u(|z|)$ is some decreasing function. The choice of $u(|z|)$ is different depending on the type of estimator. For the Krasker-Welsch approach, $u(|z|) = \xi(c/|z|)$ and

$$\xi(c/s) = (c/s)^2 + (1-(c/s)^2)(2\Phi(c/s)-1) - 2(c/s)\phi(c/s)$$

where c is a given constant, Φ represents the standard normal cumulative distribution and ϕ is the density of Φ. Hampel et al. (1986) show that a necessary condition for the existence of the transformation matrix **A**, is that the tuning constant $c > \sqrt{p}$. We can now find the *M*-estimates of covariance and location defined by the system of equations from (6.10) and

$$\sum_{i=1}^{n} w(|z_i|) z_i = 0 \tag{7.14}$$

where $z_i = \hat{\mathbf{A}}(x_i - \hat{t})$, $\hat{t}$ is an estimate of location and w is a suitable weight function. For the Krasker-Welsch (KW) approach, $w(|z|) = 1 / |z|$. So, using the Krasker-Welsch method and understanding that the distances for the *M*-estimates of covariance are $|z|$, the π-weights are the w-function, which are just the inverses of the distances.

Huber (1977, 1981) and Stahel (1981) show that the breakdown point of affine *M*-estimators of covariances is at most $1/p$, which can be a problem in large dimension data. Fortunately, in some instances these estimators work well above their breakdown point. Consider a 10-variable scenario consisting of points arranged in a fractional factorial setting with axial points. The fractional factorial is a $2^{10\text{-}4}$ design (64 points) with 16 axial points used to represent high leverage positions. The percentage of high leverage points is then 16/80 or 20%. The theoretical breakdown of the *M*-estimate of covariance is $1/p \cong 9\%$. Computing *M*-estimate (KW-type) covariance distances correctly identifies the axial points as high leverage (Table 7.1).

Table 7.1. Krasker-Welsch Distances for the 10-Variable, 20% Outlier Dataset

Case(s)	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	*M*-estimate Distance
1-64	±1	±1	±1	±1	±1	±1	±1	±1	±1	±1	17.0
65	10	0	0	0	0	0	0	0	0	0	48.9
66	0	12	0	0	0	0	0	0	0	0	58.4
67	0	0	14	0	0	0	0	0	0	0	68.2
68	0	0	0	16	0	0	0	0	0	0	77.8
69	0	0	0	0	10	0	0	0	0	0	48.9
70	0	0	0	0	0	12	0	0	0	0	58.4
71	0	0	0	0	0	0	14	0	0	0	74.7
72	0	0	0	0	0	0	0	16	0	0	85.3
73	0	0	0	0	0	0	0	0	10	0	53.5
74	0	0	0	0	0	0	0	0	0	12	64.1
75	-14	0	0	0	0	0	0	0	0	0	68.2
76	0	-16	0	0	0	0	0	0	0	0	77.8
77	0	0	-10	0	0	0	0	0	0	0	48.9
78	0	0	0	-12	0	0	0	0	0	0	58.4
79	0	0	0	0	-14	0	0	0	0	0	68.2
80	0	0	0	0	0	-16	0	0	0	0	77.8

7.3.3 The π-weights

The π-weights corresponding to the Krasker-Welsch method are computed as the inverse of the distances for each observation. This approach is simple and makes sense. Unfortunately, implementation of this technique in a GM-estimator reveals that the π-weights are in many cases large for interior **X**-space points. The resulting argument of the ψ-function is large, resulting in excessive downweighting. Both Krasker and Welsch (1982) and Walker (1984) consider the amount of overall downweighting a significant factor in the effectiveness of a GM-technique. The preferred estimator is the one that downweights only the discrepant observations. Any unnecessary downweighting is not only undesirable, but results in lower efficiency relative to least squares under normal errors.

A modification is made to the π-weights in an attempt to reduce overall downweighting and to increase efficiency. Instead of using the inverse of the distances, the numerator of the π-weight expression becomes the median distance observation. So, the π-weights become

$$\pi_i = \operatorname*{med}_j |z_j| / z_i \tag{7.15}$$

which reduces the magnitude of the π-weights significantly, while maintaining the same relative impact. The high leverage points result in π-weights less than one, increasing the argument of the ψ-function. Monte Carlo results (Chapter 5) show this to be an effective improvement over the Krasker-Welsch weights.

7.3.4 The ψ-function

Although many specific proposals for the ψ-function are available, they can all be grouped into one of two classes, monotone or redescending. The monotone methods (e.g. Huber's) continue to downweight observations of increasing argument magnitude, while not assigning zero weight to an observation unless the argument is zero. They are stable in terms of convergence properties. The redescending functions (e.g. Tukey's biweight) will assign zero weight to points with residual π-weight combinations beyond a certain point. Thus, if it is suspected that large outliers or outliers in high leverage points are present, the redescending functions will be more robust. Convergence problems can result with redescending functions, so care should be taken that a good initial estimate is obtained. Holland and Welsch (1977) and Birch (1980) state that, for redescending functions, only local convergence can be assured.

Because the Huber and Tukey ψ-functions have different strengths and because not as much is known regarding their performance in the combined problem, both functions will be used for the simulation experiments.

7.3.5 Convergence Technique

Solving the system of nonlinear equations formed by the normal equations of the bounded influence objective function requires some sort of iterative technique. Several iterative approaches have been developed including iteratively reweighted least squares (IRLS), Newton's method, and a technique proposed by Huber referred to as the H-algorithm. Each of these techniques performs well in general. IRLS is the most popular because any least squares program can be used to solve the equations and it often converges in just a few iterations. One consideration in a multi-stage robust estimator is the characteristics of the estimator after the final stage is implemented. Fully iterated convergence may not maintain some of the desirable properties of estimates from earlier stages.

Those interested in maintaining more of the properties of the initial estimate came up with the idea of limiting the number of steps of the iterative algorithm. Rousseeuw (1984) proposed this idea for IRLS, suggesting a one-step reweighted least squares (RLS) approach be used after an initial LMS estimate. The purpose of the technique is to increase the efficiency while maintaining the high breakdown characteristics of the initial estimate. One proposal tested in this paper's simulation experiment is a robust-ridge technique that uses a one-step RLS method.

7.4 Performance Comparison

Previous performance comparisons performed on robust-ridge estimators consist of evaluations by Askin and Montgomery (1984), who compare several robust-biased estimators using designed experiments with Monte Carlo simulation, and Pfaffenberger and Dielman (1990) who use an approach similar to Askin-Montgomery to compare a LAV robust estimator combined with ridge regression. LAV estimation has an unbounded objective function like *M*-estimation and does not perform well with high leverage outliers. A simulation experiment is developed for this paper to compare the Askin-Montgomery method, the Walker method, and several variations of the proposed robust-ridge estimator. The simulation uses a Monte Carlo approach so that the known coefficients can be compared to the estimates of the various techniques. The primary performance statistic used to evaluate estimation accuracy is the mean square error of estimation computed as

$$\text{MSEE} = \left(\hat{\beta}_R - \beta\right)'\left(\hat{\beta}_R - \beta\right)$$

where $\hat{\beta}_R$ is a vector of robust-ridge technique parameter estimates and β is the vector of true model coefficients. Simulation experiments are often replicated to reduce the variability in the performance statistic, so the *average* mean square error of estimation (AMSEE) is often computed.

A single designed experiment is developed to test the methods under various outlier-collinearity conditions. The factors varied include the location of the outliers in **X**-space, and the degree of multicollinearity measured using the condition number. The condition number is a diagnostic used to detect multicollinearity which assesses the degree of ill-conditioning of the **X′X** matrix. Ill-conditioning can be measured by examining the eigenvalues of **X′X,** say λ_1, λ_2, ... λ_p. One or more small eigenvalues imply the existence of near linear dependencies among the regressors. The condition number of **X′X** is

$$\kappa = \lambda_{\max} / \lambda_{\min} \tag{7.16}$$

The condition number represents the spread of the eigenvalues of **X′X**. Condition numbers less than 100 indicates no serious problem with multicollinearity, values between 100 and 1000 indicate moderate to strong collinearity, while values over 1000 are indications of severe multicollinearity. The condition numbers used as factor levels for degrees of multicollinearity in this experiment are 0, 100, and 1000.

The outlier factor in the experiment consists of three settings as well. The factor levels are zero outliers, interior **X**-space outliers, and exterior **X**-space or high leverage outliers. An outlier density of 15% is used so that for the 6-variable, 40 observation design that is used throughout, six outliers are used for the interior and exterior settings. The experiment is a two-factor, three level design with nine treatment combinations; one treatment has no outliers or collinearity, four have either outliers or collinearity, and four have some degree of both outliers and multicollinearity.

7.4.1 The Designed Experiment

The six-variable **X** matrix with various degrees of outlier-collinearity is generated from an initial orthogonal 2^{6-2} fractional factorial design. This 32-point base design consisting of ±1's, is then augmented with 8 axial points located a distance from the center that depends on the requirement for high leverage points. For the interior **X**-space location, the axial points are placed at a unit radius for the six-dimension design ($\sqrt{6}$). For the exterior **X**-space condition, the axial points are placed a distance 9.5 units from the center. The outliers are generated in the errors by drawing random variates from a heavy-tailed probability distribution. The 15% outliers are controlled to a certain degree by simulating a scaled contaminated normal distribution. The 34 nonoutliers have errors drawn from a N(0, σ^2) and the 6 outliers have errors drawn from a N(0, 100σ^2). The noise level σ^2 is set at 0.04 and the coefficient values are fixed at 4.0 for each **x** variable, resulting in a signal-to-noise ratio near 2500. The responses **y** are obtained from the linear model equation

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon$$

where **X** is the fractional factorial design with axial points, β is the vector of known coefficients and ε is the vector of random errors drawn from the scale contaminated normal distribution.

To generate the ill-conditioning effect, a method suggested by Askin and Montgomery (1984) is used to scale the columns of **X** so that the eigenvalues obtain the desired condition number (see Appendix A). The unscaled eigenvalues for each variable are (1, 3^2, 5^2, 7^2, 9^2, 10^2) for condition number = 100 and (1, 3^3, 5^3, 7^3, 9^3, 10^3) for condition number = 1000. The eigenvalues are then scaled so that they sum to the number of parameters, resulting in standardized regression coefficients. The nine different treatment combinations are replicated fifty times to increase the precision of the MSEE estimate for each technique. The result is an *average* MSEE or AMSEE computed as

$$\text{AMSEE} = \text{mean}\left[\left(\hat{\beta}_R - \beta\right)'\left(\hat{\beta}_R - \beta\right)\right] \tag{7.17}$$

The methods and datasets were developed and analyzed using the Windows version of the statistical software program S-PLUS. The software library ROBETH (Marazzi 1993) of S functions is used to aid in the computation of many of the robust regression algorithms. The code used to compute one of the proposed techniques is provided in Appendix C.

7.4.2 Simulation Results

The performance of the methods proposed by Askin and Montgomery (labeled ASKIN), and Walker (WALKER) are compared to the performance of variations of the method proposed in Section 7.3. Prior to performing the simulation runs, the development of the previously published techniques were verified using the results of the examples in their respective papers. Also included in the simulation are results from least squares estimation and robust estimation using the robust GM-method described in Section 7.3. These two techniques are added to make relative comparisons to different baselines. The variations of the proposed robust-ridge method include an approach using robust GM-estimation for the initial $\hat{\beta}$ and $\hat{\sigma}^2$, followed by ridge regression using the fully iterated Hoerl and Kennard (1976) biasing parameter k and the initial GM-weights. This technique follows the robust, then ridge philosophy and for that reason is labeled RRBIF (**R**obust **R**idge **B**ounded **I**nfluence **F**irst). All of the proposed variations in this study use the Hoerl and Kennard fully iterated method for estimating the biasing parameter. The second variation consists of using the single stage *S*-estimator robust method for the initial estimate ($\hat{\beta}$ and $\hat{\sigma}^2$), estimating k, augmenting the **X** matrix, and performing a one-step RLS bounded influence estimation. This sequence involves inserting ridge estimation between the two stages of GM-estimation, so it can be thought of as a robust-ridge-robust approach. It is labeled RROS-T to indicate **R**obust-**R**idge **O**ne-**S**tep, using the redescending **T**ukey biweight ψ-function. The final two variations modify the RROS method by iterating until convergence on the augmented GM-estimate. The two variations differ in the ψ-function used. These approaches are RRFI-T (**R**obust-**R**idge **F**ully **I**terated - **T**ukey's) and RRFI-H (**R**obust-**R**idge **F**ully **I**terated - **H**uber's). The method labeled GMP-T is the robust method baseline used for comparison. It is a robust GM-technique using an *S*-estimate initial, Tukey's biweight, and one-step RLS final stage estimation.

The estimations for all eight techniques on nine experimental combinations, each containing 50 replicates, takes approximately two hours of processing time on a 486DX2-66 PC. The numbers reported in performance comparisons are the AMSEE values computed from the 50 replicates. The AMSEE values are also rank ordered within each experimental combination, so that the sum and standard deviation of the ranks can be used as an additional evaluation tool. Because some of the rank orders can be determined by very small differences in AMSEE and others by large AMSEE differences, another measure is calculated to capture the within treatment variability. The percent over minimum AMSEE statistic is determined using the ratio of a technique's AMSEE to the best technique's AMSEE for a particular experiment treatment combination. The final statistic reported is the average mean square inefficiency ratio (AMSIR), which measures the performance of a robust-ridge or robust technique relative to least squares. Table 7.2 shows the results of the simulation experiment using the four performance measures. The results of the first run show the efficiencies of each method relative to least squares. The next two runs are pure multicollinearity problems, so that ridge effectiveness is evaluated. Runs four and seven are pure outlier problems (interior and exterior), so the robust effectiveness is compared. The other four runs, numbers five, six, eight and nine have the combined multicollinearity-outlier problem and will obviously be used to evaluate the robust-ridge estimation. Adding least squares and the bounded influence method to this study allows one to see how much improvement these combined methods can achieve over a nonrobust and purely robust approach.

In terms of efficiency, several methods including ASKIN, RRFI-T, and RRFI-H perform within 6% of least squares. The other methods have AMSIR values of 1.12 - 1.28, meaning they have AMSEE values that are 12%-28% higher than least squares. The reduced efficiency of these other methods can be attributed to either the downweighting of good points (as in WALKER) or the reduced number of iterations in final estimation (for RRBIF and RROS-T).

Table 7.2. Robust-Ridge Technique Monte Carlo Simulation Results

AMSEE

Outlier Location	Condition No.	LS	ASKIN	WALKER	RROS-T	RRBIF	RRFI-T	RRFI-H	GMP-T
None	0	0.26	0.27	0.30	0.34	0.34	0.28	0.28	0.34
None	100	2.40	2.54	2.65	3.08	3.06	2.72	2.55	3.02
None	1000	11.49	8.16	8.22	10.57	10.32	8.92	8.49	17.14
Interior	0	3.66	0.84	0.56	0.36	0.36	0.37	0.50	0.36
Interior	100	41.26	6.66	5.49	2.80	2.78	2.75	3.86	3.25
Interior	1000	203.12	15.11	14.49	11.95	13.34	11.39	13.05	25.27
Exterior	0	14.39	7.80	1.11	1.45	1.45	0.98	1.02	1.45
Exterior	100	76.69	18.99	7.55	7.79	9.06	8.49	6.61	13.06
Exterior	1000	515.35	36.61	25.89	37.53	46.74	19.25	22.54	109.67
	Sum	**868.63**	**96.99**	**66.25**	**75.87**	**87.45**	**55.15**	**58.92**	**173.56**

Relative Ranks

Outlier Location	Condition No.		ASKIN	WALKER	RROS-T	RRBIF	RRFI-T	RRFI-H	GMP-T
None	0		1	4	6	5	2	3	7
None	100		1	3	7	6	4	2	5
None	1000		1	2	6	5	4	3	7
Interior	0		7	6	3	1	4	5	2
Interior	100		7	6	3	2	1	5	4
Interior	1000		6	5	2	4	1	3	7
Exterior	0		7	3	5	4	1	2	6
Exterior	100		7	2	3	5	4	1	6
Exterior	1000		4	3	5	6	1	2	7
		Sum	**41**	**34**	**40**	**38**	**22**	**26**	**51**
		Std Deviation	**2.8**	**1.6**	**1.7**	**1.7**	**1.5**	**1.4**	**1.7**

Percent Over Minimum AMSEE

Outlier Location	Condition No.		ASKIN	WALKER	RROS-T	RRBIF	RRFI-T	RRFI-H	GMP-T
None	0		0%	10%	25%	25%	3%	3%	25%
None	100		0%	4%	21%	20%	7%	0%	19%
None	1000		0%	1%	30%	26%	9%	4%	110%
Interior	0		133%	56%	0%	0%	3%	39%	0%
Interior	100		142%	100%	2%	1%	0%	41%	18%
Interior	1000		33%	27%	5%	17%	0%	15%	122%
Exterior	0		694%	13%	47%	47%	0%	4%	48%
Exterior	100		187%	14%	18%	37%	28%	0%	97%
Exterior	1000		90%	35%	95%	143%	0%	17%	470%
	Sum		**1279%**	**259%**	**243%**	**317%**	**50%**	**123%**	**909%**

AMSIR

Outlier Location	Condition No.		ASKIN	WALKER	RROS-T	RRBIF	RRFI-T	RRFI-H	GMP-T
None	0		1.02	1.12	1.28	1.28	1.05	1.06	1.28
None	100		1.06	1.10	1.28	1.27	1.13	1.06	1.26
None	1000		0.71	0.72	0.92	0.90	0.78	0.74	1.49
Interior	0		0.23	0.15	0.10	0.10	0.10	0.14	0.10
Interior	100		0.16	0.13	0.07	0.07	0.07	0.09	0.08
Interior	1000		0.07	0.07	0.06	0.07	0.06	0.06	0.12
Exterior	0		0.54	0.08	0.10	0.10	0.07	0.07	0.10
Exterior	100		0.25	0.10	0.10	0.12	0.11	0.09	0.17
Exterior	1000		0.07	0.05	0.07	0.09	0.04	0.04	0.21
	Sum		**4.12**	**3.52**	**3.98**	**3.99**	**3.40**	**3.36**	**4.82**

The best performing techniques on the pure collinearity problem are the two previously published methods and the two fully iterated proposed methods. All involve IRLS until convergence, which appears to be an important factor in ridge estimation. The ASKIN method performs slightly better than the others, perhaps due to the efficiency of the least squares initial estimate.

Runs four through six are interior **X**-space outlier problems with varying degrees of multicollinearity. Run four has no **X'X** ill-conditioning, so the focus is solely on the robust estimation aspects. The best methods are RRBIF, RROS-T, and RRFI-T. Fortunately, these methods perform as well as the pure robust method GMP-T, meaning that the ridge estimation aspect of combined estimators does not degrade their robust performance. Runs five and six, which are combined problem scenarios, show the proposed method doing better in general than the published techniques. The two Tukey ψ-function methods, RROS-T and RRFI-T perform the best, followed by the RRBIF and RRFI-H techniques. The ASKIN and WALKER methods lag behind the proposed methods, but are considerable improvements over least squares. For the moderate multicollinearity condition (run five) the GMP-T method performs better than the ASKIN and WALKER methods, but the published methods outperform the strictly robust technique under severe multicollinearity.

Runs seven through nine are the exterior point or high leverage outlier scenarios. These runs offer the greatest challenge to all the methods, particularly those with unbounded influence functions. Even the pure robust GM-method, which is designed to be robust to the high leverage outliers, has trouble with this problem when even moderate multicollinearity is introduced. The MSEE values have large variances in some instances for these runs, often created by one or two wild estimations. Additional information can be gained by viewing plots of the mean, median and standard deviation of the MSEE results for these three runs (see Figure 7.1a, b, and c).

As expected, each of the bounded influence methods, pure or combined, performs well on run seven. The ASKIN approach performs significantly worse for this scenario and for run eight, the moderate multicollinearity problem. Differences in the AMSEE values for the five bounded influence robust-ridge methods are small and can be primarily attributed to differences in the standard deviations of the MSEE. A few aberrant estimates from the RROS-T and RRBIF inflate their AMSEE values. The same behavior is evident in run nine. The ASKIN method has an AMSEE value not too different from the RROS-T and RRBIF methods, but for different reasons. The ASKIN method has a higher median MSEE value, but smaller standard deviation caused by less deviant estimations. The WALKER, RRFI-T, and RRFI-H perform the best overall for the three exterior point runs. This result also holds for all nine runs.

Each technique's AMSIR values are plotted in Figure 7.2. The plot shows some similar behavior of certain techniques across the different scenarios. The ASKIN and GMP-T have the most trouble with the exterior point runs. The RROS-T and RRBIF methods perform in very similar fashions, for primarily the same reason. Both methods do not fully iterate on the final estimate weights. The WALKER method performs the most like the fully iterated proposed methods, but has some trouble with the interior point outlier scenarios. The two fully iterated proposed methods, RRFI-T and RRFI-H have the best overall AMSEE values. The best performing proposed technique (RRFI-T) and published technique (WALKER) will be further evaluated using an example dataset.

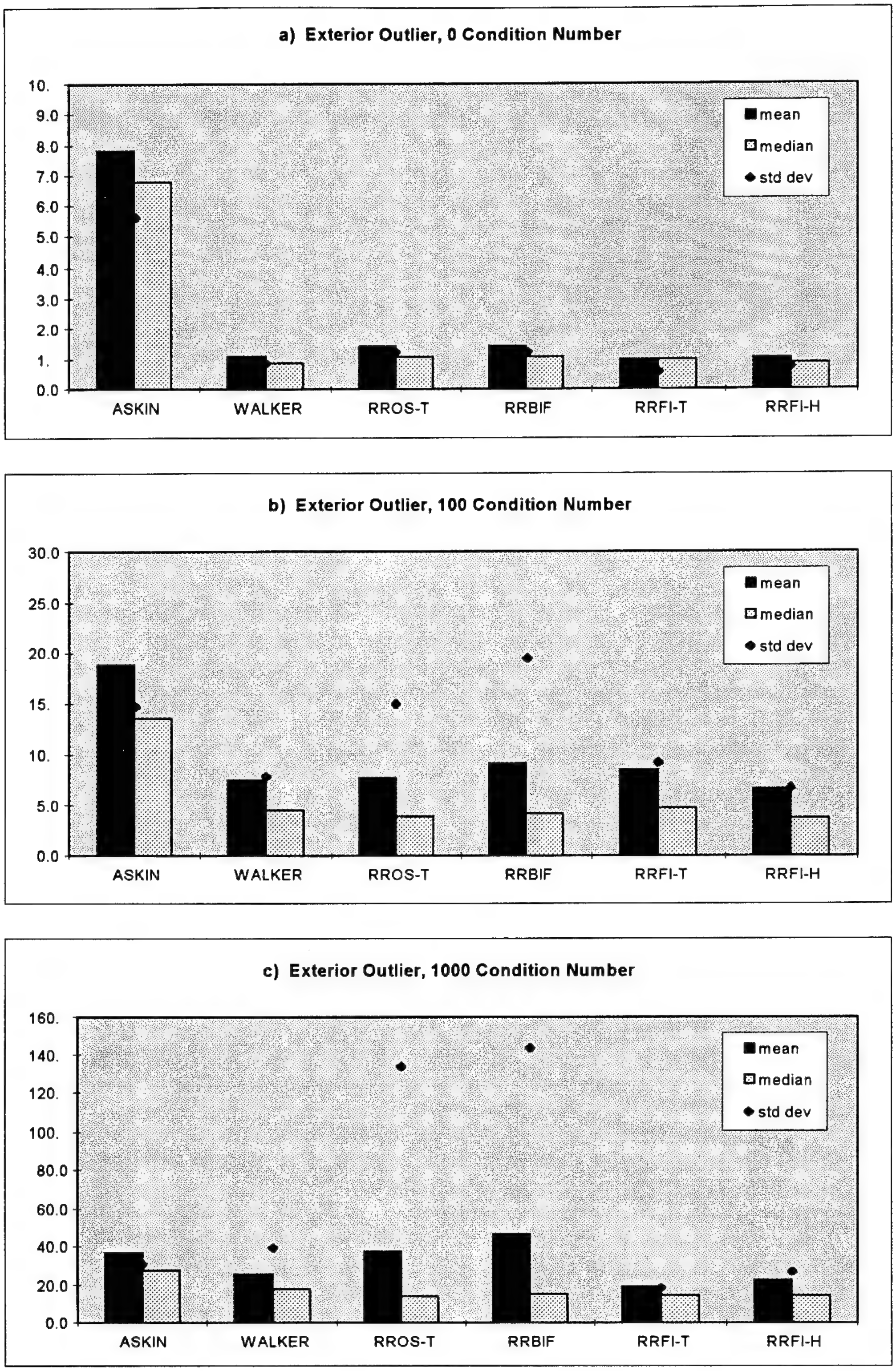


Figure 7.1. Performance of Robust-Ridge Techniques on Exterior Point Outlier Runs. a) Run 7, b) Run 8, c) Run 9.

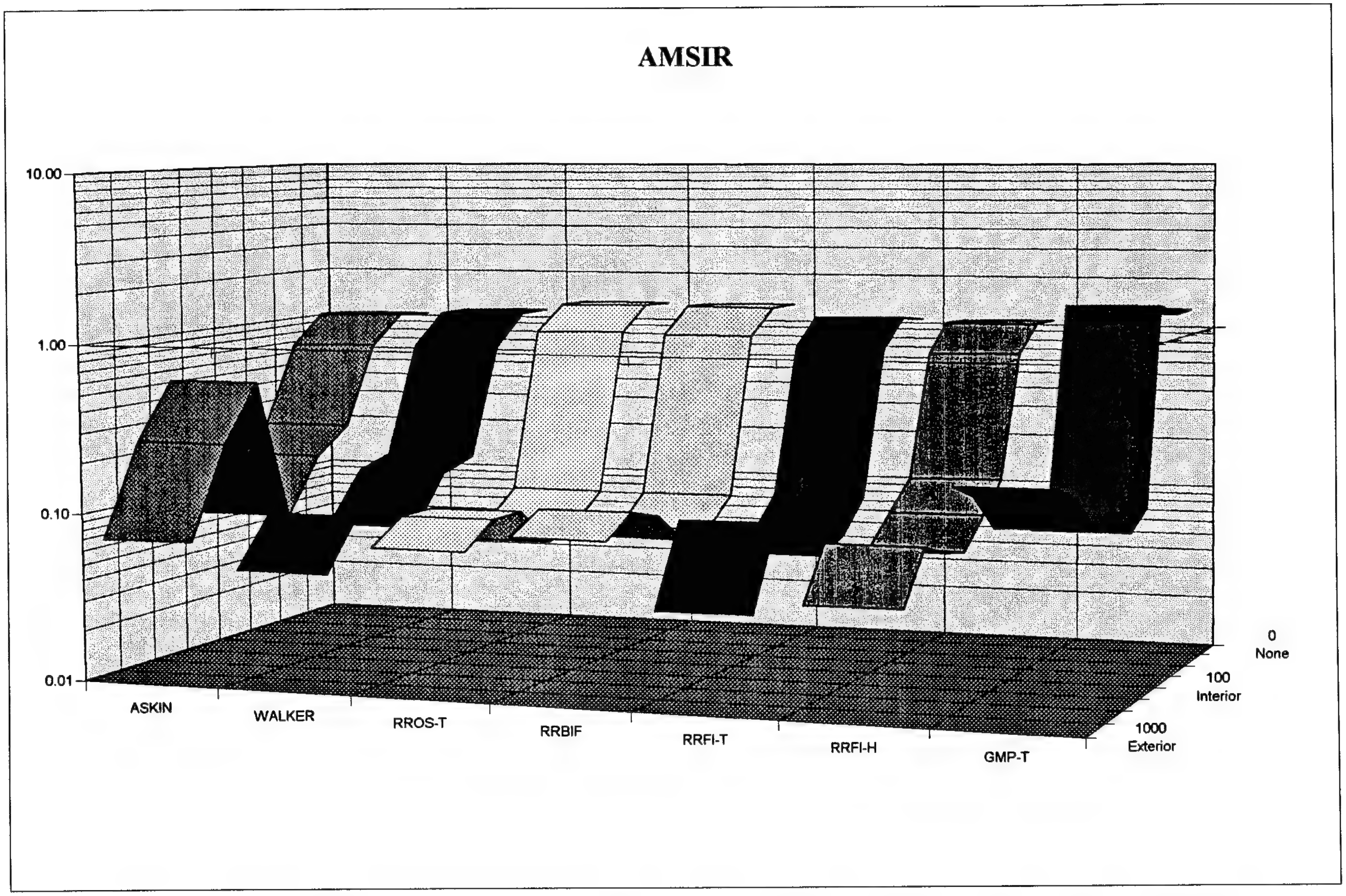


Figure 7.2. Robust-Ridge Technique Average Mean Square Inefficiency Ratio for Monte Carlo Simulation

7.5 An Example

Because not many papers have introduced robust-ridge techniques, not too many examples have been offered that adequately challenge these techniques under the combined problem. It was thus decided to take a popular and exemplary robust method dataset and modify it by adding multicollinearity to the regressor variables. The dataset of Hawkins, Bradu, and Kass (1984, Table 4) is selected because it contains outliers positioned in such a manner that makes them hard to detect. The dataset consists of three regressor variables and 75 observations, or cases. The outliers are all high leverage points (cases 1-10) located near each other in a multiple point cloud. The authors decided to further complicate the matter by including four other high leverage points (cases 11-14) located in a separate cloud which are not outliers, but are in-line with the remaining observations (cases 15-75). To generate multicollinearity, we decided to add a fourth variable that is nearly a linear combination of the first three variables. The values for the fourth variable are determined using the relation

$$\mathbf{X4} = \mathbf{X1} - \mathbf{X2} + \mathbf{X3} + \mathbf{e} \tag{7.18}$$

where **e** is a noise vector containing random variates from a standard normal distribution. The fourth variable is incorporated in the model by creating a modified response

$$\mathbf{y}' = \mathbf{y} - 0.1*(\mathbf{X4}).$$

The modified response should ensure that the added variable is also significant. The additional regressor and modified responses should not alter the outlier or leverage conditions of the original data. Because **X4** is a function of the first three variables, it will be high leverage for cases 1-14 as well. The modified response will maintain that cases 1-10 are outliers, and cases 11-75 are inliers. The multicollinearity diagnostics of this modified Hawkins, Bradu, and Kass dataset show that a severe linear dependency exists. The condition number is 1119, indicating severe multicollinearity and creating a challenge for biased estimation techniques.

Table 7.3. Robust-Ridge Coefficient Estimates for the Modified Hawkins-Bradu-Kass Dataset

		WALKER			RRFI-T	
	Robust	Ridge	Robust-Ridge	Robust	Ridge	Robust-Ridge
Intercept	-0.925	-0.926	-0.942	-0.361	-0.284	-0.135
X1	0.066	0.115	0.136	0.156	0.120	0.042
X2	0.263	0.203	0.195	0.076	0.023	0.028
X3	0.104	0.097	0.094	-0.109	-0.046	-0.041
X4	-0.016	0.024	0.032	-0.082	-0.104	-0.088

The WALKER and RRFI-T robust-ridge techniques are compared in terms of their parameter estimates (Table 7.3). In addition to showing their final estimates, we display two other sets of estimates. The first set are the technique's initial robust estimates, which includes GM-estimation using *DFFITS* for WALKER, and *S*-estimation for RRFI-T. The second set of estimates is the result of applying ridge regression without iterating on the bounded influence weights.

The purpose of evaluating the three sets of estimates for each technique is to witness the dynamics of the estimates as the techniques progress through their stages of estimation. Both techniques have parameter estimates that experience significant change during estimation. Both methods actively shrink the initial robust estimates, which is expected considering the value of the condition number. The RRFI-T final estimate of the **X4** coefficient is close to the -0.1 factor used in the modified response equation (7.18), which is encouraging. The large differences between the two techniques in both the initial and final estimates indicates that perhaps one of the methods is performing better than the other in correctly identifying outliers. A study of the final weights should provide the necessary insight.

Each technique uses a bounded influence function. Their aim is to correctly identify the high leverage points and to downweight only those high leverage points with high residuals, the "bad" leverage points. If the bounded influence functions are operating correctly in this instance, we should see final weights near zero for cases 1-10 and close to one for cases 11-14. The final weights for the other 61 observations should also be near one. The final weights for the two methods are shown Table 7.4.

Table 7.4. Final Estimate Weights for Modified H-B-K Dataset

Case	WALKER	RRFI-T
1	0.98	**0.00**
2	0.96	**0.00**
3	1.00	**0.00**
4	0.94	**0.00**
5	1.00	**0.00**
6	0.98	**0.00**
7	0.88	**0.00**
8	0.93	**0.00**
9	0.94	**0.00**
10	0.99	**0.00**
11	**0.00**	0.99
12	**0.00**	0.88
13	**0.00**	0.14
14	**0.00**	0.97
15-75	*0.98	*0.96
* mean weight for cases 15-75		

The WALKER method masks cases 1-10 and swamps cases 11-14. Hawkins, Bradu and Kass indicate that this masking/swamping characteristic also occurs for a number of robust techniques including *M*-estimation. The WALKER technique misidentifies outliers primarily because it uses inadequate measures of leverage for this data configuration. The hat diagonal measures are influenced by the multiple point clouds to such an extent that only case 14 has a high h_{ii} value (Table 7.5). The *M*-estimates of covariance (Krasker-Welsch distances) used in RRFI-T correctly identify the high leverage points so that the good leverage points can be distinguished from the bad in the data.

Table 7.5. Measures of Leverage for Modified H-B-K Dataset

Case	Hat Diag	KW Dist
1	0.07	**41.5**
2	0.06	**42.5**
3	0.09	**45.7**
4	0.08	**46.5**
5	0.07	**45.6**
6	0.08	**42.4**
7	0.07	**43.2**
8	0.06	**42.0**
9	0.12	**47.6**
10	0.09	**44.5**
11	0.10	**51.8**
12	0.15	**55.7**
13	0.11	**53.3**
14	**0.56**	**80.6**
15	0.05	12.6
* mean weight for cases 15-75		

Although correct measures of leverage are not assurances that the appropriate good and bad leverage points will be identified, it is an important step in that direction. Both of the stages of the proposed GM-estimation method, the *S*-estimation initial and bounded influence final step, correctly downweighted the first 14 cases of this dataset. This example clearly demonstrates the importance of a robust measure of leverage.

7.6 Conclusions

Many of the regression situations encountered in practical situations have a degree of linear dependency among the regressors **and** error distributions that are heavy-tailed. The techniques considered in this paper all offer a considerable improvement over either least squares, a pure biased estimation technique or a pure robust method for datasets with the simultaneous multicollinearity-outlier problem. Monte Carlo simulation of previously published and proposed methods indicates that some of the proposed methods are the overall best performing techniques. The fully iterated GM-robust-ridge methods (RRFI-T and RRFI-H) perform well in the presence of a number of combined problem scenarios. In addition, these techniques perform well when neither or only one of the problems occurs, indicating they may be used regardless of the dataset characteristics. The best performing proposed method, RRFI-T uses a Tukey ψ-function that downweights large outliers by assigning zero weight, resulting in slightly better performance than the Huber ψ-function alternative. The RRFI-T method has the lowest total AMSEE, smallest rank sum, lowest percent over minimum AMSEE, and is within 1% of the smallest sum AMSIR. Using a particularly challenging dataset with severe multicollinearity and difficult to detect outliers, the RRFI-T method properly shrinks the parameter coefficients using ridge regression and using its robust capabilities correctly identifies the high leverage inliers from the high leverage outliers. Continued work in this area is important, especially in the area of high breakdown measures of leverage. If a more stable, dependable high breakdown method were developed, it could be easily inserted in the proposed GM-method.

Chapter 8

Summary, Conclusions, and Recommended Areas for Future Research

Several related topics were considered and studied in the previous five chapters. These topics include the development of outlier datasets, the study of robust methods and the study of biased-robust methods. A brief review of the dissertation objective and the main issues in each of the chapters is provided in the following summary section. The significant results of this study are compiled and discussed in the conclusions. A final section offers some possible future research directions for those interested in working in this area.

8.1 Summary

The objective of this research is to: a) evaluate existing robust methods to determine the overall best performing methods, b) develop and evaluate new alternative robust estimators, c) perform a comprehensive evaluation of existing and new robust methods, d) develop a new biased-robust estimation technique, and e) compare existing biased-robust estimation methods with the new biased-robust technique. The approach to accomplishing this objective is to first perform robust candidate screening experiments, while at the same time determining which outlier configurations most comprehensively test the estimators' capabilities. The initial outlier configuration study consists of combinations of outlier magnitude and location used to evaluate the techniques' efficiency and bounded influence properties. The second study involves the development of various alternative GM-estimators followed by an evaluation and comparison of the competing alternatives. Several of the best performing GM-estimation techniques are then combined with the best performing existing methods for a thorough evaluation of efficiency, breakdown, resistance to multiple point outlier clouds, and performance of many outlier location/outlier density variations. The best of the existing and new methods are then studied in more detail and used to estimate costs of satellites in a real word example. The final study consists of the development of a biased-robust estimation routine based on suggestions by Askin and Montgomery (1980). The best performing robust method from this study is incorporated in the proposed biased-robust technique. This new technique has several variations which are tested against each other and against the two competing methods suggested by Askin and Montgomery (1980), and Walker (1984). An example is generated from a previously published dataset and used to test these methods.

8.2 Conclusions

The study of regression analysis methods for situations when the least squares assumptions (normal errors and independent regressors) are violated, is a topic of tremendous interest and importance for those who build models for real data. Research on methods for dealing with the outlier and multicollinearity problems increases not only our understanding of the weaknesses of current methods, but also points us in directions for alternative, improved solutions. Advances in these fields of study continue to enhance the practitioner's toolkit and will ideally reach the point when a robust, biased, or biased-robust approach will be as commonly applied as least squares. In this section we will describe the significant findings of this study in our effort to contribute to the advances in robust estimation and biased-robust estimation.

The major contributions from this dissertation to the fields of robust regression estimation and biased-robust regression estimation are:

- The development of a new multi-stage, multi-property robust regression estimator that performs as well as any robust estimator tested across a variety of nonoutlier and outlier scenarios.
- A comprehensive comparative study of the top performing proposed and existing robust regression methods in terms of their ability to accurately estimate model parameters to data without outliers and with outliers.
- The development of a new biased-robust regression estimator that performs better than previously published techniques.
- A comparative study of the new and previously published biased-robust estimators that includes a Monte Carlo simulation experiment and an application.

A number of specific conclusions relating to these contributions have been mentioned at the ends of the previous five chapters. Some of the most significant statements will be summarized in this section as a prelude to the discussion in the next section on areas for future research. The conclusions will be grouped by the major contribution areas discussed above.

8.2.1 New Robust Regression Estimator

- Many of the GM-estimation alternatives originally proposed tend to estimate poorly when outliers are present in the interior **X**-space positions only. These techniques used above average initial estimates, such as *S*-estimation or most B-robust estimation, and fully iterated convergence. The above average initial estimates on these types of datasets resulted in poor final estimates after full GM-estimation convergence. The initial estimates would properly identify and downweight the interior X-space outliers, but the GM-estimation convergence would, after several iterations, start downweighting the nonoutlying exterior X-space observations. The cause for this behavior is not known, but the solution in this case is to perform only a single GM-estimation step.

- The findings regarding the use of monotone versus redescending ψ-functions for one-step convergence GM-estimators are that the redescending functions tend to estimate slightly better because they give zero weight to extreme outliers. The monotone functions such as Huber's method maintain nonzero weights even for the extreme outliers. In some instances, these nonzero weights, although typically on the order of 10^{-3}, can have relatively significant effects on the parameter estimates. Sometimes though, it may be important for the otherwise valid observations to remain in the model with nonzero weights. This tradeoff should be understood by the analyst and it is recommended that both types of estimations be performed. The final parameter estimates and final weights should then be inspected and the information can be used to make a final method selection decision.
- The three property Coakley-Hettmansperger estimator did perform well in the experiments and showed no evidence that it lacks any of the three desirable properties. However, there was some evidence of moderate performance relative to other robust methods in areas such as tests for efficiency, and multiple point cloud experiments. The top performing GM-estimate, GMNP5, does not necessarily have high breakdown in large dimension problems, but did outperform the three property Coakley-Hettmansperger method in an overall sense. It is evident in this case that techniques with more desirable properties do not necessarily perform better.

- The analyst should be aware of the potential for poor estimates that do not fit the bulk of the data if redescending ψ-functions are used with fully iterated convergence. An example is provided in Chapter 4 in the GM alternative GMCH3. This estimator uses the Tukey ψ-function with fully IRLS. One of the datasets estimated by the technique has multiple point outliers located in a cloud that the LTS initial estimate and MVE leverage measure did not properly identify. The resulting estimates masked the outliers and swamped some of the inliers, causing the parameter estimates to be far from the true coefficients.
- The GMNP5 proposal performs better than any other proposed methods and as well as the best performing existing robust technique, MM-estimation. Either of these two methods can be applied to nearly any dataset with or without outliers and be expected to estimate a model to the majority of the data. These two methods could obviously be used in an automated regression analysis environment, where outliers are possible and exploratory data analysis is not extensively used. These techniques are also diagnostic aids because their estimates and weights can be compared to least squares to identify potential outliers in the data.

8.2.2 Comparative Study of Robust Methods

- Nearly all robust methods improve in estimation performance as outliers increase in magnitude. Estimation accuracy increases at different rates for the various methods, but they do increase as outlier magnitudes increase. Beyond a certain outlier magnitude (which depends on other characteristics of the data), robust method accuracy tends to level off. This estimation accuracy behavior is due to the insensitivity of robust methods to outliers and their ability to downweight these discrepant points. As the outliers become more and more obvious, the robust methods increase the downweighting, driving the final weights near zero, which results in increased estimation accuracy relative to the nonoutlier observations.

- High breakdown point estimators of leverage such as the minimum volume ellipsoid estimator (MVE) and the minimum covariance determinant (MCD) are not necessarily the best candidates for robust estimators of leverage, as is suggested a number of times in the literature. Some of the significant weaknesses of these approaches include extensive computation times for their exact solutions, large variability in their approximate solutions, general tendencies to swamp (identify inliers as outliers), and inabilities to identify outliers in several multiple point outlier cloud arrangements.
- The two most popular high breakdown point robust estimators, LMS and LTS, do not perform well relative to other robust estimators. These estimators have low efficiency relative to least squares. They also do not perform well against multiple point outlier clouds, interior **X**-space outliers and exterior **X**-space outliers. Other than minor differences in efficiency, these two techniques tend to perform very similarly.
- Although the literature states that Schweppe-type GM-estimators downweight high leverage points only if their corresponding residuals are large (Krasker and Welsch 1982; Hampel et al. 1986; Coakley and Hettmansperger 1993, among others), numerous simulation analyses show that this statement is not necessarily true. Many Schweppe-type estimators tested in this dissertation downweight small residual, high leverage points. For example, the GM method proposed by Marazzi (1993) downweights many of the observations in normally distributed error data (see Chapter 4). The most important determinants of a Schweppe objective's success or failure in downweighting only large residual high leverage points are the estimators of leverage, the π-weights used, the ψ-function and the associated tuning constant.

- Results of initial existing robust method experiments showed that GM-estimation and MM-estimation performed better than all the other techniques which includes *M*-estimation, LAV, LMS, and LTS techniques. Further experiments added *S*-estimation, and most B-robust estimation. GM- and MM-estimation are multiple stage, multiple property techniques, whereas all of the others are single stage, single property methods. The additional stages generate the additional properties. These multiple property approaches are the superior performing techniques and should be the focus of future research efforts.
- Robust estimators with higher breakdown do not necessarily perform better than robust methods with low breakdown. Even if the percentage of outliers is higher than a techniques breakdown point, it also does not mean that the estimator will not identify and downweight all of the outliers. For instance, *M*-estimators with $1/n$ breakdown consistently located and properly downweighted all of the outliers for datasets with outlier densities as high as 25%. Although high breakdown is a desirable property, it should be considered in relation to estimator performance.

8.2.3 New Biased-Robust Regression Estimator

- Although one-step convergence methods increase the performance capability of robust-only methods, fully iterated convergence is the better approach for biased-robust methods. The reason that fully iterated techniques improve biased-robust estimation is that the final weights are also a function of the biasing parameter. Full convergence accomplishes the needed parameter modifications resulting from both biased estimation and robust estimation.
- The sequence of biased and robust estimation for the combined estimator has been a matter of discussion in the literature. The two main arguments for initiating one approach before the

other are the comments of Belsley, Kuh and Welsch (1980) and Mason and Gunst (1985). Belsley et al. (1980, p. 210) prescribe the biased, then robust approach because "collinearity can even disguise anomalous data, ... Thus, we provisionally conclude that reduction in collinearity should be a first step for the effective detection of unusual data components." The opposite side of the issue, that of outlier inducing collinearity, is observed and reported by Mason and Gunst. They notice that when an observation that has large values on two or more predictor variables (high leverage points in two or more dimensions), collinearities can be induced. So, based on these two reports, it is unclear which estimation sequence is best. As a result, the primary methods proposed in Chapter 7 involve a three-stage approach of robust-biased-robust estimation. The robust methods are used to generate initial estimates for the biasing parameter in ridge regression. These first two steps address the Mason-Gunst concern of outlier inducing collinearity by performing initial robust estimation. Performing ridge estimation in the second stage prior to robust final estimation follows the guidance of Belsley et al. and the approaches of Askin and Montgomery (1980), and Walker (1984). These two stages provide for possible unmasking of influence by performing ridge prior to final robust estimation.

- The best performing proposed method (RRFI-T) performs better than the two previously published methods. This technique uses a robust-biased-robust approach consisting of an *S*-estimate, followed by ridge estimation, and concluding with a fully iterated GM-estimate. This methods outperforms the other two techniques in a Monte Carlo simulation study and against the Walker method using an example with outliers and multicollinearity.

8.2.4 Comparative Study of Biased-Robust Methods

- The results of the Monte Carlo study of the three biased-robust methods indicate that the primary performance difference lies in the robust estimation aspect of the combined estimators. The Askin-Montgomery method uses *M*-estimation which clearly has trouble with high leverage outliers. The Walker approach uses GM-estimation, but uses a measure for leverage, the h_{ii} , that is not robust. The modified Hawkins, Bradu and Kass dataset illustrates the weakness of the Walker approach. In general, the overall strength of the proposed RRFI-T method stems from its ability to perform well in robust estimation.
- All of the biased-robust techniques compared in this study perform better overall than least squares, robust-only, or biased-only methods on all types of problems involving varying degrees of outliers and/or multicollinearity. The augmented robust approach starting with an initial robust estimate is used in this paper. This sequence combines the biased and robust estimation techniques in such a way so that their estimations are successful against clean data, outliers only, collinearity only and the simultaneous outlier-collinearity problem. This performance indicates that these techniques could be used if the dataset characteristics are not known and good overall performance is desired.

8.3 Recommendations for Future Research

This study has made some contributions to the fields of robust and biased-robust estimation, but there is obviously much more to be learned in both of these areas that will keep researchers busy for many years ahead. Most of the suggestions provided in this section on future research topics were uncovered in the research process and may enhance the findings of this dissertation. The topics will be discussed using the sequence of the dissertation chapters.

8.3.1 Robust Methods and Outliers

Outlier Magnitude Study - It was learned from initial experiments that robust methods increased in estimation accuracy as outliers increased in error magnitude. It was also hypothesized that the improved performance is related to the increased downweighting of these outliers and increased focus on the nonoutliers in the data. A future research topic is the study of an analytical result showing this relationship between model fit and outlier magnitude.

Prior Outlier Information Study - This dissertation focuses on robust and biased-robust methods that fit well to data assuming that we have no prior reliable diagnostic information characterizing the data. The problem changes significantly if we know that the data has outliers in certain locations. For instance, we could recommend certain methods that specialize in certain outlier configurations and possibly do better than the best overall robust technique proposed in Chapter 6. If we know that the outliers are in interior **X**-space positions, *M*-estimation with the Tukey biweight ψ-function is one of the best approaches. Obviously, the research focus in this case would be the issue of dependable data diagnostics.

8.3.2 GM-estimation

GM Initial Estimate - The requirement for the GM initial estimate to be high breakdown is intuitive because it potentially adds the third desirable property to the already efficient and bounded influence approach. Unfortunately, we are not sure how high of a breakdown is needed. With most of the high breakdown methods the tuning constants can be adjusted for tradeoffs between breakdown and efficiency. Increased efficiency at the expense of unnecessary breakdown may improve overall estimation performance. Additional efficiency is especially desired on GM approaches that use limited convergence iterations. Perhaps a 25% breakdown initial *S*-estimate with a 75% efficiency would increase performance.

Estimates of Scale - Besides the *S*-estimate of scale used with an initial *S*-estimate, the primary scale estimate used in GM-estimation is the median absolute deviation (MAD), which has a simple explicit formula, high breakdown (50%) and acceptable efficiency of 37%. There are alternatives to the MAD that have similar breakdown and higher efficiency. Rousseeuw and Croux (1991) propose two estimators, S_n and Q_n ,which have gaussian asymptotic efficiencies of 58% and 82% respectively. They have developed algorithms for their computation (Croux and Rousseeuw 1992). It is unclear how well these techniques perform in general and whether their computation times are reasonable.

High Breakdown Point Estimates of Leverage - The concept of a high breakdown estimate of leverage for a GM-estimator is attractive. Simpson et al. (1992) note that for a final GM-estimate to have high breakdown, it is especially important that the measures of leverage be high breakdown. Some methods have been proposed that have high breakdown, such as the MVE and the MCD. Unfortunately, these methods have some drawbacks that inhibit their effectiveness such as computational complexity, variability in the approximate solutions, tendency to swamp inliers and inability to properly locate some multiple point cloud arrangements. More work in this area may identify solutions to these shortcomings. Rousseeuw proposes a χ^2 statistic to be used as a cutoff for the robust distances from MVE. There may be better cutoffs. Also, the MCD is faster computationally, it is asymptotically normal, and it has more stable approximations than the MVE.

However, a cutoff statistic has not been proposed. Further research on the MCD may be the most rewarding.

π-weight Study - Once a decision is made regarding which type of leverage measure to use in the GM-estimate, the next step is to determine the appropriate corresponding π-weights. Proposals exist for each of the most common measures of leverage, but their effectiveness varies. In Chapter 4 we mention that the Krasker Welsch π-weights, which are just the inverse of the distances, tend to cause too much downweighting of nonoutliers. More research on enhanced π-weight configurations is also needed.

Convergence - The tendency of the fully iterated convergence methods used in this study is to sometimes move to a worse final estimate than originally obtained in the initial estimate. This anomally is the primary reason that the one-step approaches are used and actually perform better overall. As part of the MM-estimate algorithm proposed by Yohai (1987), the final *M*-estimate convergence continues to iterate as long as improvement is observed in the objective function. This enhancement is definitely worth testing for IRLS in GM-estimation.

Inference - There has been a significant amount of work done in the area of GM-estimation inference (see Chapter 6). Perhaps these results could be used for inference, parameter estimate testing and other hypothesis tests on the proposed GM-estimates. The research could determine the asymptotics of the proposed estimator using an initial *S*-estimate and one-step reweighted least squares.

8.3.3 Biased-Robust Estimation

Alternatives to the Proposed Methods - Although several different combinations of initial estimates and final estimate approaches were implemented in this study, there may be some combinations of robust and biased methods that perform better. The only biased estimation approach used in this study was ridge regression. Although this approach has been shown to work well in previous studies (Askin and Montgomery 1984; Pfaffenberger and Dielman 1990) and also worked well in this study, there may be methods that perform even better. One alternative for selecting an optimum biasing parameter in ridge has been proposed by Lee and Campbell (1985). There may also be better performing sequences of estimation other than the robust-biased-robust approach suggested.

References

Askin, R. G., and Montgomery, D. C. (1980), "Augmented Robust Estimators," *Technometrics*, 22, 333-341.

Askin, R. G., and Montgomery, D. C. (1984), "An Analysis of Constrained Robust Regression Estimators," *Naval Research Logistics Quarterly*, 31, 283-296.

Beaton, A. E., and Tukey, J. W. (1974), "The Fitting of Power Series, Meaning Polynomials, Illustrated on Band-Spectroscopic Data," *Technometrics*, 16, 147-185.

Belsley, D. A., Kuh, E., and Welsch, R. E. (1980), *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*, Wiley: New York.

Bickel, P. J. (1973), "On Some Analogues to Linear Combination of Order Statistics in the Linear Model," *Annals of Statistics*, 1, 597-616.

Bickel, P. J. (1975), "One-step Huber Estimates in the Linear Model," *Journal of the American Statistical Association*, 70, 428-434.

Birch, J. B. (1980), "Some Convergence Properties of Iterated Reweighted Least Squares in the Location Model," *Communications in Statistics - Simulation and Computers*, B9, 359-369.

Birch, J. B., and Agard, D. B. (1993), "Robust Inference in Regression: A Comparative Study," *Communications in Statistics - Simulation*, 22, 217-244.

Boscovich, R. J., (1757) "De letteraria expeditione per pontifician ditionem, et synopsis amplioris operis ac habertur plura eius ex exemplaria etiam sensroum impressa," *Bononiensi Scientiarum et Artium Instituto Atque Academia Commentarii*, 4, 353-396.

Box, G. E. P. (1953), "Non-normality and Tests on Variances," *Biometrika*, 40, 318-335.

Butler, R. W., Davies, P. L., and Jhun, M. (1993), "Asymptotics for the Minimum Covariance Determinant Estimator," *The Annals of Statistics*, 21, 1385-1400.

Coakley, C. W. and Hettmansperger, T. P. (1993), "A Bounded Influence, High Breakdown, Efficient Regression Estimator," *Journal of the American Statistical Association,* 88, 872-880.

Cook, R. D., and Hawkins, D. M. (1990), "Comment on 'Unmasking Multivariate Outliers and Leverage Points,' by P. J. Rousseeuw and B. C. van Zomeren," *Journal of the American Statistical Association*, 85, 640-644.

Dodge, Y. (1984), "Robust Estimation of Regression Coefficients by Minimizing a Convex Combination of Least Squares and Least Absolute Deviations," *Computational Statistics Quarterly*, 139-153.

Edgeworth, F. Y. (1887), "On Observations Relating to Several Quantities," *Hermathena*, 6, 279-285.

Forsythe, A. B. (1972), "Robust Estimation of Straight Line Regression Coefficients by Minimizing p-th Power Deviations," *Technometrics*, 14, 159-166.

Gentlemen, W. M. (1965), "Robust Estimation of Multivariate Location by Minimizing p-th Power Transformations," unpublished Ph.D. dissertation, Princeton University.

Green, P. J. (1984), "Iteratively Reweighted Least Squares for Maximum Likelihood Estimation, and Some Robust and Resistant Alternatives," *Journal of the Royal Statistical Society B*, 46, 149-170.

Hampel, F. R. (1968), "Contributions to the Theory of Robust Estimation," unpublished Ph.D. dissertation, University of California, Berkley.

Hampel, F. R. (1973), "Robust Estimation: A Condensed Partial Survey," *Zeitschrift für Wahrscheinlichkeitsthiorie und Verwandte Gebiete*, 27, 87-104.

Hampel, F. R. (1974), "The Influence Curve and its Role in Robust Estimation," *Journal of the American Statistical Association*, 69, 383-393.

Hampel, F. R. (1975), "Beyond Location Parameters: Robust Concepts and Methods," *Bulletin of the International Statistical Institute*, 46, 375-382.

Hampel, F. R. (1978), "Optimally Bounding the Gross-error Sensitivity and the Influence of Position in Factor Space," *Proceedings of the Statistical Computing Section of the American Statistical Association*, Washington, D. C., 59-64.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., and Stahel, W. A. (1986), *Robust Statistics: The Approach Based on Influence Functions*, Wiley: New York.

Handshin, E., Schweppe F. C., Kohlas, J., and Fiechter, A. (1975), "Bad Data Analysis for Power System State Estimation," *IEEE Transactions on Power Apparatus Systems*, PAS-94, 329-337.

Harter, H. L. (1974-1976), "The Method of Least Squares and Some Alternatives," Parts I-VI, *International Statistical Review*, 42, 147-174 (Part I), 42, 235-264 (Part II), 43, 1-44 (Part III), 43, 125-190 (Part IV), 43, 269-278 (Part V), 44, 113-159 (Part VI).

Hawkins, D. M. (1994), "The Feasible Solution Algorithm for the Minimum Covariance Determinant Estimator in Multivariate Data," *Computational Statistics and Data Analysis*, 17, 197-210.

Hawkins, D. W., Bradu, D., and Kass, G. V. (1984), "Location of Several Outliers in Multiple-Regression Data Using Elemental Sets," *Technometrics*, 26, 197-208.

Heiler, S. (1981), "Robust Estimates in Linear Regression - A Simulation Approach" , *Computational Statistics*, eds. H. Buning, and P. Naeve, De Gruyter: Berlin, 115-136.

Hemmerle, W. J. (1975), "An Explicit Solution for Generalized Ridge Regression," *Technometrics*, 17, 309-314.

Heritier, S., and Ronchetti, E. (1994), "Robust Bounded-Influence Tests in General Parametric Models," *Journal of the American Statistical Association*, 89, 897-903.

Hettmansperger, T. P., and Sheather, S. J. (1991), "A Cautionary Note on the Method of Least Median Squares," *The American Statistician*, 46, 79-83.

Hill, R. W. (1977), "Robust Regression When There Are Outliers in the Carriers," unpublished Ph.D. dissertation, Harvard University, Boston, MA.

Hoaglin, D. C., and Welsch, R. E. (1978), "The Hat Matrix in Regression and ANOVA," *American Statistician*, 32, 17-22.

Hoerl, A. E., and Kennard, R. W. (1970a), "Ridge Regression: Biased Estimation for Non-orthogonal Problems," *Technometrics*, 12, 55-67.

Hoerl, A. E., and Kennard, R. W. (1970b), "Ridge Regression: Applications to Non-orthogonal Problems," *Technometrics*, 12, 69-82.

Hoerl, A. E., and Kennard, R. W. (1976), "Ridge Regression: Iterative Estimation of the Biasing Parameter," *Communications in Statistics*, A5, 77-88.

Hoerl, A. E., Kennard, R. W., and Baldwin, K. F. (1975), "Ridge Regression: Some Simulations," *Communications in Statistics*, 4, 105-123.

Hogg, R. V. (1979), "An Introduction to Robust Estimation," *Robustness in Statistics*, eds. R. Launer and G. Wilkinson, Academic Press: New York, 1-17.

Holland, P. W. (1973), "Weighted Ridge Regression: Combining Ridge and Robust Regression Methods," NBER Working Paper Series, Working Paper #11, 1-19.

Holland, P. W. and Welsch, Roy E. (1977), "Robust Regression Using Iteratively Reweighted Least-Squares," *Communications in Statistics - Theory and Methods*, A6, 813-827.

Huber, P. J. (1964), "Robust Estimation of a Location Parameter," *Annals of Mathematical Statistics*, 35, 73-101.

Huber, P. J. (1965), "A Robust Version of the Probability Ratio Test," *Annals of Mathematical Statistics*, 36, 1753-1758.

Huber, P. J. (1972), "Robust Statistics: A Review," *Annals of Mathematical Statistics*, 43, 1041-1067.

Huber, P. J. (1973), "Robust Regression: Asymptotics, Conjectures, and Monte Carlo," *The Annals of Statistics*, 1, 799-821.

Huber, P. J. (1977), "Robust Covariances," *Statistical Decision Theory and Related Topics*, Vol. 2, eds. Gupta and Moore, Academic Press: New York, 165-191.

Huber, P. J. (1981), *Robust Statistics*, Wiley: New York.

Huber, P. J. (1993), "Projection Pursuit and Robustness," *New Directions in Statistical Data Analysis and Robustness*, eds. Morgenthaler, Ronchetti, and Stahel, Birkhäuser Verlag: Basel, Switzerland, 139-146.

Jaeckel, L. A. (1972), "Estimating Regression Coefficients by Minimizing the Dispersion of Residuals," *Annals of Mathematical Statistics*, 5, 1449-1458.

Jureckova, J. (1971), "Nonparametric Estimate of Regression Coefficients," *Annals of Mathematical Statistics*, 42, 1328-1338.

Koenker, R., and Bassett, G. J. (1978), "Regression Quantiles," *Econometrica*, 46, 33-50.

Krasker, W. S. (1980), "Estimation in Linear Regression Models with Disparate Data Points," *Econometrica*, 48, 1333-1346.

Krasker, W. S., and Welsh, R. E. (1982), "Efficient Bounded-Influence Regression Estimation," *Journal of the American Statistical Association*, 77, 595-603.

Lawrence, K. D., and Marsh, L. C. (1984), "Robust Ridge Estimation Methods for Predicting U. S. Coal Mining Fatalities," *Communications in Statistics-Theory and Methods*, 13, 139-149.

Lee, T., and Campbell, D. B. (1985), "Selecting the Optimum K in Ridge Regression," *Communications in Statistics-Theory and Methods*, 14, 1589-1604.

Lopuhaa, H. P., and Rousseeuw, P. J. (1991), "Breakdown Points of Affine Equivariant Estimators of Multivariate Location and Covariance Matrices," *The Annals of Statistics*, 19, 229-248.

Mallows, C. L. (1973), "Some Comments on Cp," *Technometrics*, 15, 661-675.

Mallows, C. L. (1975), "On Some Topics in Robustness," unpublished memorandum, Bell Telephone Laboratories: Murrray Hill, NJ.

Marazzi, A. (1993), *Algorithms, Routines, and S Functions for Robust Statistics*, Wadsworth and Brooks/Cole: Pacific Grove, California.

Markatou, M. and He, X. (1994) "Bounded Influence and High Breakdown Point Testing Procedures in Linear Models," *Journal of the American Statistical Association*, 89, 543-559.

Markatou, M., and Hettmansperger, T. P. (1990), "Robust Bounded-Influence Tests in Linear Models," *Journal of the American Statistical Association*, 85, 187-190.

Maronna R., and Yohai, V. J. (1991), "The Breakdown Point of Simultaneous General M Estimates of Regression and Scale," *Journal of the American Statistical Association*, 86, 699-703.

Maronna, R. A. (1976), "Robust *M*-estimators of Multivariate Location and Scatter," *Annals of Statistics*, 4, 51-67.

Maronna, R. A., and Yohai, V. J. (1981), "Asymptotic Behavior of General M-estimates for Regression and Scale with Random Carriers," *Zeitschrift für Wahrscheinlichkeitsthiorie und Verwandte Gebiete*, 58, 7-20.

Maronna, R. A., Yohai, V. J., and Zamar, R. J. (1993), "Bias-Robust Regression Estimation: A Partial Survey," *New Directions in Statistical Data Analysis and Robustness*, eds. S. Morganthaler, E. Ronchetti, and W. Stahel, Birkhäuser Verlag: Basel, Switzerland, 157-176.

Marquardt, D. W. (1970), "Generalized Inverses, Ridge Regression, Biased Linear Estimation, and Nonlinear Estimation," *Technometrics*, 12, 591-612.

Mason, R. L., and Gunst, R. F. (1985), "Outlier-Induced Collinearities," *Technometrics*, 27, 401-407.

McDonald, G. C., and Galarneau, D. I. (1975), "A Monte Carlo Evaluation of Some Ridge-type Estimators," *Journal of the American Statistical Association*, 70, 407-416.

Montgomery D. C., and Askin, R. G. (1981), "Problems of Nonnormality and Multicollinearity for Forecasting Methods Based on Least Squares," *AIIE Transactions*, 13, 102-115.

Montgomery, D. C., and Friedman, D. J. (1993), "Prediction Using Regression Models with Multicollinear Predictor Variables," *IIE Trransactions*, 25, 73-85.

Montgomery, D. C., and Peck, E. A. (1992), *Introduction to Linear Regression Analysis* (2nd ed.), Wiley: NewYork.

Pariente, S., and Welsch, R. E. (1977), "Ridge and Robust Regression Using Parametric Linear Programming," Working Paper, MIT, Alfred P. Sloan School of Management.

Pfaffenberger, R. C., and Dielman, T. E. (1985), "A Comparison of Robust Ridge Estimators," *Business Economics Section Proceedings of the American Statistical Association*, 631-635.

Pfaffenberger, R. C., and Dielman, T. E. (1990), "A Comparison of Regression Estimators When Both Multicollinearity and Outliers Are Present," *Robust Regression: Analysis and Applications*, eds. K. Lawrence and J. Arthur, 243-270.

Rey, W. J. J. (1983), *Introduction to Robust and Quasi-Robust Statistical Methods*, Springer-Verlag: Berlin, Germany.

Ronchetti, E. (1982), "Robust Testing in Linear Models: The Infinitesimal Approach," unpublished Ph.D. dissertation, ETH, Zurich.

Ronchetti, E. (1987), "Bounded Influence Inference in Regression: A Review," *Statistical Data Analysis Based on the L1-norm and Related Methods*, ed. Y. Dodge, 65-80.

Rousseeuw, P. J, and Yohai, V. J. (1984), "Robust Regression by Means of *S*-Estimators," *Robust and Nonlinear Time Series Analysis*, eds. J. Franke, W. Hardle, and D. Martin, Springer-Verlag: Heidelberg, Germany, 256-272.

Rousseeuw, P. J. (1983), "Multivariate Estimation with High Breakdown Point," *Mathematical Statistics and Applications,* Vol. B, eds. W. Grossmann, G. Pflug, I. Vincze, and W. Wertz, Reidel: Dordrecht, The Netherlands, 283-297.

Rousseeuw, P. J. (1984), "Least Median of Squares Regression," *Journal of the American Statistical Association*, 79, 871-880.

Rousseeuw, P. J., and Croux, C. (1991), "Alternatives to the Median Absolute Deviation," Technical Report 91-43, Universitaire Instelling Antwerpen, Belgium.

Rousseeuw, P. J., and Croux, C. (1993), "Alternatives to the Median Absolute Deviation," *Journal of the American Statistical Association*, 88, 424, 1273-1283.

Rousseeuw, P. J., and Leroy, A. M. (1987), *Robust Regression and Outlier Detection*, Wiley: NewYork.

Rousseeuw, P. J., and van Zomeren, B. C. (1990), "Unmasking Multivariate Outliers and Leverage Points," *Journal of the American Statistical Association*, 85, 633-651.

Rousseeuw, P. J., and van Zomeren, B. C. (1991), "Robust Distances: Simulations and Cutoff Values," *Directions in Robust Statistics and Diagnostics,* Part II, eds. W. Stahel and S. Weisberg, Springer-Verlag: Heidelberg, Germany, 195-203.

Ruppert, D. (1992), "Computing *S* Estimators for Regression and Multivariate Location/Dispersion," *Journal of Computational and Graphical Statistics*, 1, 3, 253-270.

Ruppert, D., and Carroll, R. J. (1980), "Trimmed Least Squares Estimation in the Linear Model," *Journal of the American Statistical Association*, 75, 828-838.

S-Plus (1994), S-PLUS Reference Manual, Statistical Sciences, Inc.: Seattle.

Simonoff, J. S. (1991), "General Approaches to Stepwise Identification of Unusual Values in Data Analysis," *Directions in Robust Statistics and Diagnostics,* Part II, eds. W. Stahel and S. Weisberg, Springer-Verlag: Heidelberg, Germany, 223-242.

Simpson, D. G., Ruppert, D., and Carroll, R. J. (1992), "On One-Step GM Estimates and Stability of Influences in Linear Regression," *Journal of the American Statistical Association*, 87, 439-450.

Sposito, V. A., Kennedy, W. J., and Gentle, J. E. (1977), "Lp norm Fit of a Straight Line," *Applied Statistics*, 26, 114-116.

Stahel, W. A. (1981), "Breakdown of Covariance Estimators," Research Report 31, Fachgruppe für Statistik, ETH, Zurich.

Stigler, S. M. (1973), "Simon Newcomb, Percy Daniell, and the History of Robust Estimation 1885-1920," *Journal of the American Statistical Association*, 68, 872-879.

Street, J. O., Carroll, R. J., and Ruppert, D. (1988), "A Note on Computing Robust Regression Estimates Via Iteratively Reweighted Least Squares," *The American Statistician*, 42, 152-154.

Tukey, J. W. (1960), "A Survey of Sampling from Contaminated Distributions," *Contributions to Probability and Statistics*, ed. Olkin, Stanford University Press: Stanford, CA, 448-485.

Walker, E. (1984), "Influence, Collinearity, and Robust Estimation in Regression," unpublished Ph.D. dissertation, Department of Statistics, Virginia Polytechnic Institute.

Walker, E. (1987), "Augmented Bounded-Influence Estimators," *Statistical Computing Section Proceedings of the American Statistical Association*, 158-164.

Walker, E., and Birch, J. B. (1985), "Influence and Collinearity" *Statistical Computing Section Proceedings of the American Statistical Association*, 113-118.

Walker, E., and Birch, J. B. (1988), "Influence Measures in Ridge Regression," *Technometrics*, 30, 221-227.

Welsch, R. E. (1977), "Regression Sensitivity Analysis and Bounded-Influence Estimation," *Evaluation of Econometric Models*, eds. Kmenta and Ramsay, Academic Press: New York, 153-167.

Yohai, V. J. (1987), "High Breakdown-Point and High Efficiency Robust Estimates for Regression," *The Annals of Statistics*, 15, 642-656.

Yohai, V. J., Stahel, W. A., and Zamar, R. H. (1991), "A Procedure for Robust Estimation and Inference in Linear Regression," *Directions in Robust Statistics and Diagnostics,* Part II, eds. W. Stahel and S. Weisberg, Springer-Verlag: Heidelberg, Germany, 365-374.

Appendix A

Generating Datasets with Multicollinearity and Outliers

Generating Datasets with Multicollinearity and Outliers

The datasets used in the Monte Carlo studies for biased-robust regression estimation can have varying degrees of multicollinearity and nonnormal error data. To control the dependency among the regressors and the percentage and magnitude of outliers, datasets are manipulated in terms of the elements of the regressor matrix and the error vector to generate the desired characteristics. Once the regressor or **X** matrix has the degree of multicollinearity desired and the errors have the necessary outliers, the response vector is generated using the linear model relation

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

where **y** is a n x 1 vector of responses; **X** is an n x p matrix of the levels of the regressor variables; β is a p x 1 vector of the model coefficients; and ε is an n x 1 vector of errors. The coefficients for the β vector are determined based on a desired signal-to-noise ratio for the model. For example, if the error variance is 1 (noise), and the desired signal-to-noise ratio is 100:1, then the model coefficients for a 2-variable model are about 7 ($7^2 + 7^2 = 98$). The error vector can be generated using draws from a heavy-tailed distribution such as the contaminated normal where

$$\varepsilon = \alpha \, \mathrm{N}(0,1) + (1 - \alpha)\mathrm{N}(0,10)$$

where α is the percentage of the distribution that is not contaminated. Further control can be placed over the errors by fixing the percentage of observations drawn for each portion of the distribution. If the desired percentage of outliers is 15% in a 20 observation sample, 17 observations can have N(0,1) errors and 3 observations have N(0,10) errors. The characteristics of a dataset that can vary for these simulation studies are explained in the following section. The subsequent section details the development of an example dataset with multicollinearity and outliers.

Factors Considered:

Regressor Dimension - Modify the number of model parameters (*p*) to determine the effect different sized models have on the estimation technique effectiveness

Sample Size - The sample size (*n*) is determined as a function of the number of model regressors. A *p* to *n* ratio is specified and used consistently for all simulations.

Exterior X-space or High Leverage Points - The degree to which points exist in the outer region of the **X**-space. These points, called high leverage points, may or may not be influential. High leverage points present particular challenges to robust methods because these techniques must be able to bound the influence of these points, especially if the random errors are large.

Regressor Multicollinearity - Multicollinearity is measured using the eigenvalues of the **X′X** matrix so that shrinkage methods can be evaluated

Error Term Distribution - The error values determine whether points are outliers. The outlier magnitude is controlled to a certain extent by determining the probability distribution for the random variate draws.

An Example

The basic approach, as suggested by Askin and Montgomery (1984), is to start with a matrix of orthogonal regressors using a factorial or fractional factorial representation of ±1's. This matrix is then augmented with points located on the regressor space axes, called axial points. The axial points can be placed any distance from the design center, providing the flexibility to generate high leverage points. Typically these augmented observations represent 10-25% of the dataset. In this example, a 2-variable model with a sample size of 14 is generated, the initial **X** matrix will be the factorial design matrix in 12 points, with the remaining two positions as high leverage axial points. If leverage points are not required, the axial points are located at a unit radius distance ($\sqrt{p}$) from the design center. In this case though, the high leverage points are located 10 units from the design center. The axial points maintain the independence of the columns (regressor variables), which provides control over the multicollinearity portion of the model. These datasets can be easily developed using S-PLUS.

$$\begin{bmatrix} -1 & -1 \\ 1 & -1 \\ -1 & 1 \\ 1 & 1 \\ -1 & -1 \\ 1 & -1 \\ -1 & 1 \\ 1 & 1 \\ -1 & -1 \\ 1 & -1 \\ -1 & 1 \\ 1 & 1 \end{bmatrix}$$

At this point the **regressor dimension** is determined. The next step is to modify the matrix so that some **high leverage points** are present. The design matrix is augmented by adding axial points that result in points that extend the **X**-space region. An example of the augmented matrix is

$$\begin{bmatrix} -1 & -1 \\ 1 & -1 \\ -1 & 1 \\ 1 & 1 \\ -1 & -1 \\ 1 & -1 \\ -1 & 1 \\ 1 & 1 \\ -1 & -1 \\ 1 & -1 \\ -1 & 1 \\ 1 & 1 \\ 0 & 10 \\ 10 & 0 \end{bmatrix}$$

To modify the design matrix for the desired eigenvalue spread, the matrix requires scaling of the columns (or regressors) of **X**. Assume the desired spread is 1:10. The existing **X′X** matrix is

$$\begin{bmatrix} 112 & 0 \\ 0 & 112 \end{bmatrix}$$

Because the **X′X** is diagonal, the eigenvalues are equal to the values of the diagonal. It is also important to transform the **X** matrix so that the regression coefficients are standardized. To do this, the sum of the eigenvalues must be equal to the rank of the matrix, which in this case is the number of parameters. So we must keep the desired ratio and have the eigenvalues sum to 2.0. The resulting desired eigenvalues are (0.182, 1.82). The eigenvalues are also equal to the sum of the squares of the columns of **X**. If the desired eigenvalues are 0.182 and 1.82, the elements in each column can be scaled so their corresponding Σx_i^2 is equal to the desired spread. For this example the new design matrix would have the following elements

$$\begin{bmatrix} -0.04 & -0.13 \\ 0.04 & -0.13 \\ -0.04 & 0.13 \\ 0.04 & 0.13 \\ -0.04 & -0.13 \\ 0.04 & -0.13 \\ -0.04 & 0.13 \\ 0.04 & 0.13 \\ -0.04 & -0.13 \\ 0.04 & -0.13 \\ -0.04 & 0.13 \\ 0.04 & 0.13 \\ 0 & 1.3 \\ 0.4 & 0 \end{bmatrix}$$

so that **X′X** would be

$$\begin{bmatrix} 0.182 & 0 \\ 0 & 1.82 \end{bmatrix}$$

and the eigenvalues are 0.182 and 1.82.

Once the expected values of the responses are obtained using the model coefficients, the error term can be introduced by drawing random variates from a particular heavy-tailed probability distribution. If the desired distribution is the scale contaminated Normal, a random variate is generated for each observation in the sample using the following approach:

- Determine a contamination ratio (e.g. 20%) and a contamination level (e.g. 10)
- Draw a uniform random number, u
- If $u \leq 0.8$ draw an error term from N(0,1)
- If $u > 0.8$ draw the term from N(0,10)

To exercise slightly more control over the number of outliers, determine the number of outliers for the data (e.g. 2). Then, if it is desired to locate the outliers in the interior **X**-space, draw from the N(0,10) for the first two design points. If high leverage outliers are required, reserve the high variance variates for the final two observations. For example, we generate high leverage outliers for this example and obtain the following error vector

$$\begin{bmatrix} -0.3 \\ 1.0 \\ 0.0 \\ 0.7 \\ -0.8 \\ 1.7 \\ 0.2 \\ 1.8 \\ 0.2 \\ -0.1 \\ 0.4 \\ -0.5 \\ 2.4 \\ -5.0 \end{bmatrix}$$

The dataset responses can now be generated using the linear equation $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \varepsilon$, so

$$\begin{bmatrix} -1.5 \\ 1.6 \\ 0.6 \\ 1.9 \\ -1.9 \\ 2.3 \\ -0.4 \\ 3.0 \\ -0.9 \\ 0.5 \\ -0.2 \\ 0.7 \\ 11.4 \\ -2.2 \end{bmatrix} = \begin{bmatrix} -0.04 & -0.13 \\ 0.04 & -0.13 \\ -0.04 & 0.13 \\ 0.04 & 0.13 \\ -0.04 & -0.13 \\ 0.04 & -0.13 \\ -0.04 & 0.13 \\ 0.04 & 0.13 \\ -0.04 & -0.13 \\ 0.04 & -0.13 \\ -0.04 & 0.13 \\ 0.04 & 0.13 \\ 0 & 1.3 \\ 0.4 & 0 \end{bmatrix} \begin{bmatrix} 7 \\ 7 \end{bmatrix} + \begin{bmatrix} -0.3 \\ 1.0 \\ 0.0 \\ 0.7 \\ -0.8 \\ 1.7 \\ 0.2 \\ 1.8 \\ 0.2 \\ -0.1 \\ 0.4 \\ -0.5 \\ 2.4 \\ -5.0 \end{bmatrix}$$

The response vector and **X** matrix can now be used in the simulation experiment for the specified treatment combination. Another way of expressing an eigenvalue spread of 1:10 is to divide the largest eigenvalue by the smallest and the result is called the condition number. So this example represents a multicollinearity condition number of 10. The errors are configured such that the dataset has high leverage outliers with an outlier density of 2/14=14%.

Appendix B

Program Code for Proposed GM-estimation Function (GMNP5 or GMP-T)

Program Code for Proposed GM-estimation Function (GMNP5 or GMP-T)

```
#  This code is used to perform the proposed GM-estimation method described below.  A regression
#  function is generated from this code that can be applied against any dataset.  This function
#  bijs5sa can also be used in the simulation studies discussed in this paper.
#
#  METHOD GMNP5
#
#
#  GM-estimation (bounded influence) of the Schweppe type
#
#
#       Initial Fit:  S-estimator (Rousseeuw and Yohai, 1984; see Marazzi (1993) p. 216)
#             Scale:  S-estimate of scale
#          Leverage:  Modified (Normalized) M-estimate of covariance (Marazzi, 1993)
#         Pi-weight:  M-estimate of leverage over the median M-estimate of leverage
#               Psi:  Tukey's Biweight
#   Tuning Constant:  c = 4.685 Tukey's for 95% efficiency
#       Convergence:  One iteration of IRLS
#

bijs5sa <-  function(x, y, w = rep(1, nrow(x)), int = TRUE, init = sest(x,y),
                     method = wt.bibisquare, wx, iter = 1,
                     acc = 50 * .Machine$single.eps^0.5, test.vec = "resid")

{
        if(int) {
                coef <- init$coef
                coefin <- coef
                x <- cbind(1, x)
        }
        else {
               init <- sest(x,y,int=FALSE)
               coefin <- coef  <- init$coef
               x <- as.matrix(x)
        }

        if(!missing(wx)) {
                if(length(wx) != nrow(x))
                        stop("Length of wx must equal number of observations")
                if(any(wx < 0))
                        stop("Negative wx value")
                w <- w * wx
        }
        if(ncol(x) != length(coef))
                stop("Must have same number of initial values as coefficients")
        resid <- y - x %*% coef

#  Determine the tuning constant based on the suggestion of Marazzi and Joss (1993)
        tc_4.685

        if (int==F)   xwt_as.matrix(cbind(1,x))
          else
        xwt_as.matrix(x)

# Robeth pi weights using the scatter matrix
        dfrpar(xwt, "Kra-Wel")
        # Weights
         z      <- wimedv(xwt)
         z      <- wynalg(xwt, z$a); nitw <- z$nit

#  Scale the distances such that the median distance is unity and all others are a ratio of the
#  actual distance to the median distance

#  If any of the design points are at the design center (z$dist=0)
```

```
      if(any(z$dist <= 1)) {
         for (i in 1:length(y))  {
           if (z$dist[i] <= 1)   z$dist[i] <- 1
         }
      }

         z$distm <- z$dist/median(z$dist)
         pi    <- 1/z$distm

#  S-estimator scale estimate
        scale <- init$smin

        for(iiter in 1:iter) {

                if(scale == 0) {
                        convi <- 0
                        method.exit <- TRUE
                        status <- "could not compute scale of residuals"
                }
                else {

                        epis_c(resid/(scale*pi))

#  In case the residuals go to zero, keeps the weight = 1 (vs undefined)
                        if(any(resid == 0)) {
                                for (i in 1:length(y)) {
                                        if (resid[i] == 0)
                                          w[i] <- 1
                                else
                                          w[i] <- method(epis[i],tc)
                                }
                        }
                        else
                                w <- method(epis,tc)
                        if(!missing(wx))
                                w <- w * wx
                        temp <- lsfit(x, y, w, int = FALSE)
                        coef <- temp$coef
                        resid <- temp$residuals
                    }
        }
        if(!missing(wx)) {
                tmp <- (wx != 0)
                w[tmp] <- w[tmp]/wx[tmp]
        }
        list(coef = coef, initialest = coefin, residuals = resid, scale = scale,
             tc = tc, distances = z$dist,
              piweight = pi, erroverpis = epis, w = w, int = int)
}

wt.bibisquare_

#  bounded influence WEIGHT FUNCTION where w(t) = psi (t) / t  and t = e / pi*s
#  The Bisquare psi function
#  user supplied tuning constant

function(u, tc=4.685)
{
        U <- abs(u/tc)
        si <- u*(1 - (u/tc)^2)^2
        si[U > 1] <- 0
        w <- si/u
        w
}
```

Appendix C

Program Code for Proposed Biased-Robust Estimation Function RRFI-T

Program Code for Proposed Biased-Robust Estimation Function RRFI-T

```
#  This code is used to perform the proposed biased-robust method described below.  A regression
#  SPLUS function is generated from this code that can be applied against any dataset. This function
#  rrjs3s can also be used in the simulation studies discussed in this paper.
#
#
#  METHOD RRFI-T
#
#  Combined ROBUST-RIDGE-ROBUST estimation using Robust-Ridge-Bounded Influence Estimation

#   Using a WEIGHTED centering and scaling routine (see Walker 1984, p. 123-124)
#
#
#       Initial Fit:  S-estimator (Rousseeuw and Yohai, 1984; see Marazzi and Joss (1993) p. 216)
#             Scale:  S-estimate of scale
#             Ridge:  Hoerl and Kennard fully iterated
#          Leverage:  KW M-estimate of covariance (Marazzi and Joss, 1993)
#                     using a Collinearity Insensitive Routine (WFCOL)
#         Pi-weight:  M-estimate of leverage over the median M-estimate of leverage
#               Psi:  Tukey's Biweight
#   Tuning Constant:  c = 4.685 Tukey's for 95% efficiency
#       Convergence:  Fully iterated IRLS
#
#
rrjs3s <- function(x, y, w = rep(1, nrow(x)), int = TRUE, method = wt.bibisquare,
                   wx, augment = TRUE, iter = 20, acc = 50 * .Machine$single.eps^0.5,
                   test.vec = "resid", standcoef = TRUE)
{

x_as.matrix(x)  # in case x is a one dimensional list (single regressor)
origx_as.matrix(x)
p_ncol(x)
n_nrow(x)

# Compute the INITIAL Estimates
# determine the biasing parameter k

  if (int==F) {
        init <- sest(x, y, int=F)
        beta1_init$coef
         w_init$w
       }

  if (int==T) {
        init <- sest(x, y)
        beta1_init$coef
        w_init$w
       }

#  Scale the Weighted X matrix


#
#  For model diagnostic purposes, convert the X matrix to unit length and scale the y vector
#  The result is a correlation matrix form of X and a set of standardized regression coefficients
#  with no intercept term
#

#
#  Weighted Unit LENGTH scaling (Walker, 1984, p. 123)
#

if (standcoef) {
```

```
xj_rep(1,p)
sj_rep(1,p)
xs_matrix(nrow=length(y),ncol=p)
k_ncol(x)
     for (j in 1:k) {
       xj[j]_(sum(w*x[,j])/sum(w))
       sj[j]_(sum(w*(x[,j]-xj[j])^2)^0.5)
       xs[,j]_(x[,j]-xj[j])/sj[j]
       }
    if (int) {
       xs_cbind(1,xs)
       p_ncol(xs)
    }
}


#
#  Multicollinearity Detection
#

#  Standardize X so that it is in correlation form

sjj_rep(1,k)
avgx_rep(1,k)
xcorr_matrix(ncol=k,nrow=length(y))

 for (i in 1:k) {
  avgx[i]_mean(x[,i])
  sjj[i]_sum((x[,i]-avgx[i])^2)
  xcorr[,i]_(x[,i]-avgx[i])/sqrt(sjj[i])
  }

xpx_t(xcorr)%*%xcorr
#  Solve for the Variance Inflation Factors (VIFs), which are
#  the diagonals of (X'X)^-1

vif_diag(solve(xpx))

#  Find the condition number (max / min eigenvalue)
eigenval_eigen(xpx)$values
lmax_max(eigen(xpx)$values)
lmin_min(eigen(xpx)$values)
condnum_lmax/lmin

#  Scale the responses (y's) to find the standardized regression coefficients
#  Use unit length scaling for the y's (M&P p. 156, Syy)

# if (standcoef) {
#  syy_(n-1)*var(y)

#  avgy_mean(y)
#  y_(y-avgy)/sqrt(syy)
# }

#  Compute the scaled coefficients and estimate of scale

if (standcoef) {
  lsout_lsfit(xs,y,w,int=F)
  betas_lsout$coef
  sscale_mad(lsout$residuals)

}

#  Can convert back to the original coefficients
#  Compute the transformation matrix


if (standcoef) {
  as_diag(1/sj,ncol=k)
  as1_cbind(0,as)
  atop_c(1,-xj/sj)
 if (int)
  A_matrix(rbind(atop,as1),ncol=p)
  else A_matrix(as)

#  Can convert back to the original coefficients
  origs_A%*%betas
}



#  S-estimator squared scale estimate for variance estimate

  if (standcoef==F)
        msestd <- init$smin^2
    else
        msestd <- sscale^2

#  If we insist on using nonstandardized X and Beta (as in Monte Carlo
  if (standcoef==F)  betas_beta1
#
#  ESTIMATING THE BIASING PARAMETER FOR RIDGE REGRESSION
#
#  Determine the value of the biasing parameter k
#  use the MSE and coefficient estimates of the LS fit
#
#  Initial estimate of k using Hoerl, Kennard, and Baldwin (1975)
#


khkb_c((p*msestd)/(t(betas)%*%betas))   # the c() function makes k a scalar
#
#  Then we can iterate on k as suggested by Hoerl and Kennard (1976)
#

knew_khkb
kold_0.000001
yaug_c(y,0*c(1:p))
tterm_sum(1/eigen(xpx)$values)/p
termval_20*(tterm)^-1.3

while ((knew-kold)/kold>termval) {
  kold_knew
  aug_sqrt(knew)*diag(nrow=p,ncol=p)
  xaug_rbind(xs,aug)
  beta_lsfit(xaug,yaug,int=F)$coef
  knew_c((p*msestd)/(t(beta)%*%beta))
  }
khk_knew

#  Augment the x and y matrices using the biasing constant k

kused_khk
n_n+p
aug_sqrt(kused)*diag(nrow=p,ncol=p)
xaug_rbind(xs,aug)
yaug_c(y,0*c(1:p))



#  Perform RIDGE regression given a value of the biasing constant
#  using the augmented matrix approach
w_c(w,rep(1,p))
ridout_lsfit(xaug,yaug,w,int=F)
coefhk_ridout$coef
```

```
#  Convert Back to the Original Coefficients

if (standcoef) {
  origrs_A%*%coefhk
}
#  ROBUST - RIDGE ESTIMATION
#
#  Perform fully iterated BI estimation using the augmented matrices
#
#


        if(ncol(xaug) != length(betas))
                stop("Must have same number of initial values as coefficients")


        irls.delta <- function(old, new)
        {
                a <- sum((old - new)^2)
                b <- sum(old^2)
                if(b >= 1 || a < b * .Machine$double.xmax)
                        sqrt(a/b)
                else .Machine$double.xmax
        }
        irls.rrxwr <- function(xaug, w, r)
        {
                w <- sqrt(w)
                max(abs((as.vector(r * w) %*% x)/sqrt(as.vector(w) %*% (x^2))))/sqrt(sum(w * r^2))
        }
        if(!(any(test.vec == c("resid", "coef", "w", "NULL")) || is.null(test.vec)))
                stop("invalid testvec")



        if(!missing(wx)) {
                if(length(wx) != nrow(xaug))
                        stop("Length of wx must equal number of observations")
                if(any(wx < 0))
                        stop("Negative wx value")
                w <- w * wx
        }
        resid <- ridout$residuals
#       resid <- yaug - xaug %*% betas
        converged <- FALSE
        status <- "converged"
        conv <- NULL
        method.in.control <- method.exit <- FALSE

#  Determine the tuning constant based on the suggestion of Marazzi and Joss (1993)
        tc_4.685

#  Robeth pi weights using the scatter matrix
        if (int) {
          xpi_as.matrix(cbind(1,x))
        }
        else
          xpi_as.matrix(x)

          dfrpar(xpi, "Kra-Wel")
          # Weights
           z      <- wimedv(xpi)
           z      <- wynalg(xpi, z$a); nitw <- z$nit


#  Scale the distances such that the median distance is unity and all others are a ratio of the
#  actual distance to the median distance
```

```
           z$distm <- z$dist/median(z$dist)

           pi    <- 1/z$distm
#  augment the pi weights for the augmented observations
           pi    <- c(pi,rep(1,p))


#  Use the scale estimate from the initial S-estimate
        scale  <- sqrt(msestd)


        for(iiter in 1:iter) {
                 if(!is.null(test.vec))
                         previous <- get(test.vec)


                 if(scale == 0) {
                         convi <- 0
                         method.exit <- TRUE
                         status <- "could not compute scale of residuals"
                 }
                 else {


                         epis_resid/(scale*pi)

#  In case the residuals go to zero, keeps the weight = 1 (vs undefined)

                         if(any(resid == 0)) {
                                 for (i in 1:length(yaug)) {
                                         if (resid[i] == 0)
                                           w[i] <- 1
                                 else
                                           w[i] <- method(epis[i],tc)
                                 }
                         }
                         else {
#
#  Keep the weights of the augmented observations = 1
#
                                 if(augment) {
                                         for (i in 1:n)
                                            if (i<=(n-p))
                                              w[i] <- method(epis[i],tc)
                                            else w[i] <- 1    # maintain w[i]=1 for augmented
#                                                               observations
                                         }

                                 else
                                         w <- method(epis,tc)

                         epis <- c(epis)
                         }

                         if(!missing(wx))
                                 w <- w * wx

#  Every time the weights change, update the scaled weighted X matrix
#  before computing the new scaled coefficients
#
     if (standcoef) {
        k_ncol(x)
        xjn_rep(1,k)
        sjn_rep(1,k)
        xs_matrix(nrow=length(y),ncol=k)
     for (j in 1:k) {
       wx_w[1:length(y)]
       xjn[j]_(sum(wx*x[,j])/sum(wx))
```

```
        sjn[j]_(sum(wx*(x[,j]-xjn[j])^2)^0.5)
        xs[,j]_(x[,j]-xjn[j])/sjn[j]
        }
     if (int) {
        xs_cbind(1,xs)
        p_ncol(xs)
     }


    xaug_rbind(xs,aug)

    }


                        temp <- lsfit(xaug, yaug, w, int = FALSE)
                        coef <- temp$coef
                        resid <- temp$residuals
                        if(!is.null(test.vec))
                                convi <- irls.delta(previous, get(test.vec))
                        else convi <- irls.rrxwr(xaug, w, resid)
                }
                conv <- c(conv, convi)
                converged <- convi <= acc
                done <- method.exit || (converged && !method.in.control)
                if(done)
                        break
        }
        if(!done)
                warning(status <- paste("failed to converge in", iter, "steps"))
        if(!missing(wx)) {
                tmp <- (wx != 0)
                w[tmp] <- w[tmp]/wx[tmp]
        }

#  Convert Back to the Original Coefficients

#  Can convert back to the original coefficients
#  Compute the transformation matrix


if (standcoef) {
  as_diag(1/sjn,ncol=k)
  as1_cbind(0,as)
  atop_c(1,-xjn/sjn)
 if (int)
  A_matrix(rbind(atop,as1),ncol=p)
  else A_matrix(as)

#  Can convert back to the original coefficients
  origrr_A%*%coef


# Compute final residuals and MSE
  if (int)
   x1_cbind(1,origx)
    else x1_as.matrix(origx)
  finres_c(y-x1%*%origrr)
  finmse_sum(finres^2)/(n-p-1)

}
  else {
  finres_resid
  finmse_sum(resid^2)/(n-p-1)
  }
```

```
if (standcoef) {
        list(finalcoef = origrr, finalstd = coef, finmse =finmse, finres = finres,
            initialstd = betas,  initial = beta1,  intialchk = origs,
            vifs = vif, conditnum = condnum,
            kinit = khkb, kiter = khk,  kused = kused,
            ridgstd = coefhk, ridgorig = origrs,
            scale = scale, tc = tc, distances = z$dist,
            w = w)
}
else
        list(finalcoef = coef, ridge = coefhk, initial = beta1,
            finmse =finmse, finres = finres,
            vifs = vif, conditnum = condnum,
            kinit = khkb, kiter = khk,  kused = kused,
            scale = scale, tc = tc, distances = z$dist,
            w = w)

}
```

Appendix D

Program Code for Simulation Experiment on Robust Regression Estimators

Program Code for Simulation Experiment on Robust Regression Estimators

```
#
#  Pilot Study for Chapter 5  - Robust-Only Monte Carlo Simulation Study of Robust Techniques
#
#  This is an example of batch file written in S-PLUS used to perform a simulation study of robust
#  techniques versus different data configurations of outlier location, outlier magnitude, number of
#  regressors, and outlier density.  Eleven regression methods are evaluated including least squares
#  for this 24 treatment design.  50 replicates of each treatment combination are run and an average
#  MSEE is computed for each technique.  The AMSEEs are stored in files for future analysis.
#
#
#  Purpose is to determine which robust technique performs the best under a variety of scenarios
#       A mixed level 3 factor outlier test - 4 x 3 x 2 experiment
#               1)  outlier location/leverage content (4 levels)
#                       a)  interior outliers and no leverage points
#                       b)  interior outliers and leverage points
#                       c)  exterior outliers and leverage points
#                       d)  interior and exterior outliers and leverage points
#               2)  number of independent variables (2 levels)
#                      a)  2 -  2 regressor, 12 design space points + 4 axial points to be used for
#                          leverage
#                      b)  6 -  6 regressor, 32 design space points + 8 axial points for leverage
#                      c) 10 - 10 regressor, 64 design space points + 16 axial points also for
#                         leverage
#               3)  outlier density (2 levels)
#                       a)  10% outliers
#                       b)  20% outliers
#**************************************************************************************************
#***                                                                                          ****
#***                    OUTLIER MAGNITUDE / OUTLIER LOCATION EXPERIMENT                       ****
#***                                                                                          ****
#**************************************************************************************************
#
#  Number of independent variables (not including the intercept)
#  desdim        the dimension of the X matrix
#            =   2 -  2 regressor, 12 design space points + 4 axial points to be used for leverage
#            =   6 -  6 regressor, 32 design space points + 8 axial points for leverage
#            =  10 - 10 regressor, 64 design space points + 16 axial points also for leverage
#
#
#  orthcoef      the values of the orthogonal coefficients
#            =   1 -  b for each regressor (such that the signal term is equal among different
dimension models)
#                     (2var=7, 6var=4, and 10var=3)  signal is the sum of the squares of the
coefficients
#

#
#  ****************************************************************************
#     X matrices for the OUTLIER LOCATION / LEVERAGE CONTENT experiments
#  ****************************************************************************

#

#  1111111111111111111111111111111111111111111111111111111111111111111111
#
#  Dataset 1 -  2 Variable, n=16, NO High Leverage Points
#               Efficiency Test,
#              3 replicates of a 2 factor full factorial + 4 axial points

x2ax1_matrix(c(-1,-1,
1,-1,
-1,1,
1,1,
-1,-1,
1,-1,
-1,1,
1,1,
-1,-1,
1,-1,
-1,1,
1,1,
1.414,0,
0,1.414,
-1.414,0,
0,-1.414),ncol=2,byrow=T)
#  111111111111111111111111111111111111111111111111111111111111111111111111


#  222222222222222222222222222222222222222222222222222222222222222222222222
#
#  Dataset 2 -  10 Variable, n=80, NO High Leverage Points,
#                Efficiency Test
#                1/16 fraction of a 10 factor factorial (64 obs) + 16 axial points

x10ax1_matrix(c(-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,
1,1,1,-1,-1,-1,-1,-1,-1,-1,
-1,1,-1,1,1,1,-1,-1,-1,-1,
1,-1,1,1,1,1,-1,-1,-1,-1,
1,-1,-1,-1,1,-1,1,-1,-1,-1,
-1,1,1,-1,1,-1,1,-1,-1,-1,
1,1,-1,1,-1,1,1,-1,-1,-1,
-1,-1,1,1,-1,1,1,-1,-1,-1,
-1,-1,-1,1,1,-1,-1,1,-1,-1,
1,1,1,1,1,-1,-1,1,-1,-1,
-1,1,-1,-1,-1,1,-1,1,-1,-1,
1,-1,1,-1,-1,1,-1,1,-1,-1,
1,-1,-1,1,-1,-1,1,1,-1,-1,
-1,1,1,1,-1,-1,1,1,-1,-1,
1,1,-1,-1,1,1,1,1,-1,-1,
-1,-1,1,-1,1,1,1,1,-1,-1,
1,1,-1,-1,-1,-1,-1,-1,1,-1,
-1,-1,1,-1,-1,-1,-1,-1,1,-1,
1,-1,-1,1,1,1,-1,-1,1,-1,
-1,1,1,1,1,1,-1,-1,1,-1,
-1,1,-1,-1,1,-1,1,-1,1,-1,
1,-1,1,-1,1,-1,1,-1,1,-1,
-1,-1,-1,1,-1,1,1,-1,1,-1,
1,1,1,1,-1,1,1,-1,1,-1,
1,1,-1,1,1,-1,-1,1,1,-1,
-1,-1,1,1,1,-1,-1,1,1,-1,
1,-1,-1,-1,-1,1,-1,1,1,-1,
-1,1,1,-1,-1,1,-1,1,1,-1,
-1,1,-1,1,-1,-1,1,1,1,-1,
1,-1,1,1,-1,-1,1,1,1,-1,
-1,-1,-1,-1,1,1,1,1,1,-1,
1,1,1,-1,1,1,1,1,1,-1,
-1,-1,-1,1,-1,-1,-1,-1,-1,1,
1,1,1,1,-1,-1,-1,-1,-1,1,
-1,1,-1,-1,1,1,-1,-1,-1,1,
1,-1,1,-1,1,1,-1,-1,-1,1,
1,-1,-1,1,1,-1,1,-1,-1,1,
-1,1,1,1,1,-1,1,-1,-1,1,
1,1,-1,-1,-1,1,1,-1,-1,1,
-1,-1,1,-1,-1,1,1,-1,-1,1,
-1,-1,-1,-1,1,-1,-1,1,-1,1,
1,1,1,-1,1,-1,-1,1,-1,1,
-1,1,-1,1,-1,1,-1,1,-1,1,
1,-1,1,1,-1,1,-1,1,-1,1,
1,-1,-1,-1,-1,-1,1,1,-1,1,
-1,1,1,-1,-1,-1,1,1,-1,1,
1,1,-1,1,1,1,1,1,-1,1,
-1,-1,1,1,1,1,1,1,-1,1,
1,1,-1,1,-1,-1,-1,-1,1,1,
-1,-1,1,1,-1,-1,-1,-1,1,1,
1,-1,-1,-1,1,1,-1,-1,1,1,
-1,1,1,-1,1,1,-1,-1,1,1,
-1,1,-1,1,1,-1,1,-1,1,1,
1,-1,1,1,1,-1,1,-1,1,1,
-1,-1,-1,-1,-1,1,1,-1,1,1,
1,1,1,-1,-1,1,1,-1,1,1,
1,1,-1,-1,1,-1,-1,1,1,1,
-1,-1,1,-1,1,-1,-1,1,1,1,
1,-1,-1,1,-1,1,-1,1,1,1,
-1,1,1,1,-1,1,-1,1,1,1,
-1,1,-1,-1,-1,-1,1,1,1,1,
1,-1,1,-1,-1,-1,1,1,1,1,
-1,-1,-1,1,1,1,1,1,1,1,
1,1,1,1,1,1,1,1,1,1,
3.162,0,0,0,0,0,0,0,0,0,
0,3.162,0,0,0,0,0,0,0,0,
0,0,3.162,0,0,0,0,0,0,0,
0,0,0,3.162,0,0,0,0,0,0,
0,0,0,0,3.162,0,0,0,0,0,
0,0,0,0,0,3.162,0,0,0,0,
0,0,0,0,0,0,3.162,0,0,0,
0,0,0,0,0,0,0,3.162,0,0,
0,0,0,0,0,0,0,0,3.162,0,
0,0,0,0,0,0,0,0,0,3.162,
-3.162,0,0,0,0,0,0,0,0,0,
0,-3.162,0,0,0,0,0,0,0,0,
0,0,-3.162,0,0,0,0,0,0,0,
0,0,0,-3.162,0,0,0,0,0,0,
0,0,0,0,-3.162,0,0,0,0,0,
0,0,0,0,0,-3.162,0,0,0,0),ncol=10,byrow=T)
#  2222222222222222222222222222222222222222222222222222222222222222222222

#  3333333333333333333333333333333333333333333333333333333333333333333333
#
#  Dataset 3 -  2 Variable, n=16, NO High Leverage Points
#                Efficiency Test,
#               3 replicates of a 2 factor full factorial + 4 axial points

x2axv_matrix(c(-1,-1,
1,-1,
-1,1,
1,1,
-1,-1,
1,-1,
-1,1,
1,1,
-1,-1,
1,-1,
-1,1,
1,1,
4,0,
0,5,
-6,0,
0,-7),ncol=2,byrow=T)
#  3333333333333333333333333333333333333333333333333333333333333333333333


#  4444444444444444444444444444444444444444444444444444444444444444444444
#
#  Dataset 4 -  10 Variable, n=80, High Leverage Points,
#                Efficiency Test
#                1/16 fraction of a 10 factor factorial (64 obs) + 16 axial points

x10axv_matrix(c(-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,
1,1,1,-1,-1,-1,-1,-1,-1,-1,
-1,1,-1,1,1,1,-1,-1,-1,-1,
1,-1,1,1,1,1,-1,-1,-1,-1,
1,-1,-1,-1,1,-1,1,-1,-1,-1,
-1,1,1,-1,1,-1,1,-1,-1,-1,
1,1,-1,1,-1,1,1,-1,-1,-1,
-1,-1,1,1,-1,1,1,-1,-1,-1,
-1,-1,-1,1,1,-1,-1,1,-1,-1,
1,1,1,1,1,-1,-1,1,-1,-1,
-1,1,-1,-1,-1,1,-1,1,-1,-1,
1,-1,1,-1,-1,1,-1,1,-1,-1,
1,-1,-1,1,-1,-1,1,1,-1,-1,
-1,1,1,1,-1,-1,1,1,-1,-1,
1,1,-1,-1,1,1,1,1,-1,-1,
-1,-1,1,-1,1,1,1,1,-1,-1,
1,1,-1,-1,-1,-1,-1,-1,1,-1,
-1,-1,1,-1,-1,-1,-1,-1,1,-1,
1,-1,-1,1,1,1,-1,-1,1,-1,
-1,1,1,1,1,1,-1,-1,1,-1,
-1,1,-1,-1,1,-1,1,-1,1,-1,
1,-1,1,-1,1,-1,1,-1,1,-1,
-1,-1,-1,1,-1,1,1,-1,1,-1,
1,1,1,1,-1,1,1,-1,1,-1,
1,1,-1,1,1,-1,-1,1,1,-1,
-1,-1,1,1,1,-1,-1,1,1,-1,
1,-1,-1,-1,-1,1,-1,1,1,-1,
-1,1,1,-1,-1,1,-1,1,1,-1,
-1,1,-1,1,-1,-1,1,1,1,-1,
1,-1,1,1,-1,-1,1,1,1,-1,
-1,-1,-1,-1,1,1,1,1,1,-1,
1,1,1,-1,1,1,1,1,1,-1,
-1,-1,-1,1,-1,-1,-1,-1,-1,1,
1,1,1,1,-1,-1,-1,-1,-1,1,
-1,1,-1,-1,1,1,-1,-1,-1,1,
1,-1,1,-1,1,1,-1,-1,-1,1,
1,-1,-1,1,1,-1,1,-1,-1,1,
-1,1,1,1,1,-1,1,-1,-1,1,
1,1,-1,-1,-1,1,1,-1,-1,1,
-1,-1,1,-1,-1,1,1,-1,-1,1,
-1,-1,-1,-1,1,-1,-1,1,-1,1,
1,1,1,-1,1,-1,-1,1,-1,1,
-1,1,-1,1,-1,1,-1,1,-1,1,
1,-1,1,1,-1,1,-1,1,-1,1,
1,-1,-1,-1,-1,-1,1,1,-1,1,
-1,1,1,-1,-1,-1,1,1,-1,1,
1,1,-1,1,1,1,1,1,-1,1,
-1,-1,1,1,1,1,1,1,-1,1,
1,1,-1,1,-1,-1,-1,-1,1,1,
-1,-1,1,1,-1,-1,-1,-1,1,1,
1,-1,-1,-1,1,1,-1,-1,1,1,
-1,1,1,-1,1,1,-1,-1,1,1,
-1,1,-1,1,1,-1,1,-1,1,1,
1,-1,1,1,1,-1,1,-1,1,1,
-1,-1,-1,-1,-1,1,1,-1,1,1,
1,1,1,-1,-1,1,1,-1,1,1,
1,1,-1,-1,1,-1,-1,1,1,1,
-1,-1,1,-1,1,-1,-1,1,1,1,
1,-1,-1,1,-1,1,-1,1,1,1,
-1,1,1,1,-1,1,-1,1,1,1,
-1,1,-1,-1,-1,-1,1,1,1,1,
1,-1,1,-1,-1,-1,1,1,1,1,
-1,-1,-1,1,1,1,1,1,1,1,
1,1,1,1,1,1,1,1,1,1,
10,0,0,0,0,0,0,0,0,0,
0,12,0,0,0,0,0,0,0,0,
0,0,14,0,0,0,0,0,0,0,
0,0,0,16,0,0,0,0,0,0,
0,0,0,0,10,0,0,0,0,0,
0,0,0,0,0,12,0,0,0,0,
0,0,0,0,0,0,14,0,0,0,
0,0,0,0,0,0,0,16,0,0,
0,0,0,0,0,0,0,0,10,0,
0,0,0,0,0,0,0,0,0,12,
-14,0,0,0,0,0,0,0,0,0,
0,-16,0,0,0,0,0,0,0,0,
0,0,-10,0,0,0,0,0,0,0,
0,0,0,-12,0,0,0,0,0,0,
0,0,0,0,-14,0,0,0,0,0,
0,0,0,0,0,-16,0,0,0,0),ncol=10,byrow=T)
#  44444444444444444444444444444444444444444444444444444444444444

# 5555555555555555555555555555555555555555555555555555555555555555
#
#  Dataset 5 -  6 Variable, n=40, NO high leverage points,
#               Efficiency Test and Low Breakdown Test,
#               1/2 fraction of a 6 factor factorial (32 obs) + 8 axial points at sqrt(p)
#
x6ax1_matrix(c(-1,-1,-1,-1,-1,-1,
1,1,-1,-1,-1,-1,
1,-1,1,-1,-1,-1,
-1,1,1,-1,-1,-1,
1,-1,-1,1,-1,-1,
-1,1,-1,1,-1,-1,
-1,-1,1,1,-1,-1,
1,1,1,1,-1,-1,
1,-1,-1,-1,1,-1,
-1,1,-1,-1,1,-1,
-1,-1,1,-1,1,-1,
1,1,1,-1,1,-1,
-1,-1,-1,1,1,-1,
1,1,-1,1,1,-1,
1,-1,1,1,1,-1,
-1,1,1,1,1,-1,
1,-1,-1,-1,-1,1,
-1,1,-1,-1,-1,1,
-1,-1,1,-1,-1,1,
1,1,1,-1,-1,1,
-1,-1,-1,1,-1,1,
1,1,-1,1,-1,1,
1,-1,1,1,-1,1,
-1,1,1,1,-1,1,
-1,-1,-1,-1,1,1,
1,1,-1,-1,1,1,
1,-1,1,-1,1,1,
-1,1,1,-1,1,1,
1,-1,-1,1,1,1,
-1,1,-1,1,1,1,
-1,-1,1,1,1,1,
1,1,1,1,1,1,
2.45,0,0,0,0,0,
0,2.45,0,0,0,0,
0,0,2.45,0,0,0,
0,0,0,2.45,0,0,
0,0,0,0,2.45,0,
0,0,0,0,0,2.45,
-2.45,0,0,0,0,0,
0,-2.45,0,0,0,0),ncol=6,byrow=T)
# 5555555555555555555555555555555555555555555555555555555555555555

#  6666666666666666666666666666666666666666666666666666666666666666
#
#  Dataset 6 -  6 Variable, n=40, 20% (8) Variable High Leverage Points,
#               Bounded Influence and High Breakdown Test
#               1/2 fraction of a 6 factor factorial (32 obs) + 8 axial points
#
x6axv_matrix(c(-1,-1,-1,-1,-1,-1,
1,1,-1,-1,-1,-1,
1,-1,1,-1,-1,-1,
-1,1,1,-1,-1,-1,
1,-1,-1,1,-1,-1,
-1,1,-1,1,-1,-1,
-1,-1,1,1,-1,-1,
1,1,1,1,-1,-1,
1,-1,-1,-1,1,-1,
-1,1,-1,-1,1,-1,
-1,-1,1,-1,1,-1,
1,1,1,-1,1,-1,
-1,-1,-1,1,1,-1,
1,1,-1,1,1,-1,
1,-1,1,1,1,-1,
-1,1,1,1,1,-1,
1,-1,-1,-1,-1,1,
-1,1,-1,-1,-1,1,
-1,-1,1,-1,-1,1,
1,1,1,-1,-1,1,
-1,-1,-1,1,-1,1,
1,1,-1,1,-1,1,
1,-1,1,1,-1,1,
-1,1,1,1,-1,1,
-1,-1,-1,-1,1,1,
1,1,-1,-1,1,1,
1,-1,1,-1,1,1,
-1,1,1,-1,1,1,
1,-1,-1,1,1,1,
-1,1,-1,1,1,1,
-1,-1,1,1,1,1,
1,1,1,1,1,1,
7,0,0,0,0,0,
0,9,0,0,0,0,
0,0,11,0,0,0,
0,0,0,13,0,0,
0,0,0,0,7,0,
0,0,0,0,0,9,
-11,0,0,0,0,0,
0,-13,0,0,0,0),ncol=6,byrow=T)
#  66666666666666666666666666666666666666666666666666666666666666666666666666666666666666666


#*********************************************************************************************
#  Measures of MVE distance for each dataset (robust squared distances)
#
#  Using Marazzi's ROBETH software which computes Rousseeuw's MVE approximation
#     using random subsamples of size n=10,000 for 6 and 10 variable datasets and
#     complete searches for 2 variable datasets


#*************************************************************************************************

mveds1_mymvlm(x2ax1,rep(1,16),ilm=0,iopt=3)$d^2
mveds2_mymvlm(x10ax1,rep(1,80),ilm=0,iopt=2,nrep=10000,iseed=12345)$d^2
mveds3_mymvlm(x2axv,rep(1,16),ilm=0,iopt=3)$d^2
mveds4_mymvlm(x10axv,rep(1,80),ilm=0,iopt=2,nrep=10000,iseed=12345)$d^2
mveds5_mymvlm(x6ax1,rep(1,40),ilm=0,iopt=2,nrep=10000,iseed=12345)$d^2
mveds6_mymvlm(x6axv,rep(1,40),ilm=0,iopt=2,nrep=10000,iseed=12345)$d^2

#*************************************************************************************************
#
#  Compute M-estimates of Covariance for the Krasker-Welsch Weights (zds#)
#


#  Set initial values for the ROBETH functions

dfvals()
```

```
# ds1
         x2ax1wi   <- cbind(1,x2ax1)
         dfrpar(x2ax1wi, "Kra-Wel")
         zds1      <- wimedv(x2ax1wi)
         zds1      <- wynalg(x2ax1wi, zds1$a)$dist

# ds2
         x10ax1wi   <- cbind(1,x10ax1)
         dfrpar(x10ax1wi, "Kra-Wel")
         zds2      <- wimedv(x10ax1wi)
         zds2      <- wynalg(x10ax1wi, zds2$a)$dist

# ds3
         x2axvwi   <- cbind(1,x2axv)
         dfrpar(x2axvwi, "Kra-Wel")
         zds3      <- wimedv(x2axvwi)
         zds3      <- wynalg(x2axvwi, zds3$a)$dist

# ds4
         x10axvwi   <- cbind(1,x10axv)
         dfrpar(x10axvwi, "Kra-Wel")
         zds4      <- wimedv(x10axvwi)
         zds4      <- wynalg(x10axvwi, zds4$a)$dist

# ds5
         x6ax1wi   <- cbind(1,x6ax1)
         dfrpar(x6ax1wi, "Kra-Wel")
         zds5      <- wimedv(x6ax1wi)
         zds5      <- wynalg(x6ax1wi, zds5$a)$dist

# ds6
         x6axvwi   <- cbind(1,x6axv)
         dfrpar(x6axvwi, "Kra-Wel")
         zds6      <- wimedv(x6axvwi)
         zds6      <- wynalg(x6axvwi, zds6$a)$dist



#  Number of replicates per design point

nrun_50


#  Dataset Looping
#

for (k in 1:24)  {
if (k==1) {desdim_2; n_16; p_3; x_x2ax1; mverd_mveds1; mdist_zds1}
if (k==2) {desdim_10; n_80; p_11; x_x10ax1; mverd_mveds2; mdist_zds2}
if (k==3) {desdim_2; n_16; p_3; x_x2ax1; mverd_mveds1; mdist_zds1}
if (k==4) {desdim_10; n_80; p_11; x_x10ax1; mverd_mveds2; mdist_zds2}
if (k==5) {desdim_2; n_16; p_3; x_x2axv; mverd_mveds3; mdist_zds3}
if (k==6) {desdim_10; n_80; p_11; x_x10axv; mverd_mveds4; mdist_zds4}
if (k==7) {desdim_2; n_16; p_3; x_x2axv; mverd_mveds3; mdist_zds3}
if (k==8) {desdim_10; n_80; p_11; x_x10axv; mverd_mveds4; mdist_zds4}
if (k==9) {desdim_2; n_16; p_3; x_x2axv; mverd_mveds3; mdist_zds3}
if (k==10) {desdim_10; n_80; p_11; x_x10axv; mverd_mveds4; mdist_zds4}
if (k==11) {desdim_2; n_16; p_3; x_x2axv; mverd_mveds3; mdist_zds3}
if (k==12) {desdim_10; n_80; p_11; x_x10axv; mverd_mveds4; mdist_zds4}
if (k==13) {desdim_2; n_16; p_3; x_x2axv; mverd_mveds3; mdist_zds3}
if (k==14) {desdim_10; n_80; p_11; x_x10axv; mverd_mveds4; mdist_zds4}
if (k==15) {desdim_2; n_16; p_3; x_x2axv; mverd_mveds3; mdist_zds3}
if (k==16) {desdim_10; n_80; p_11; x_x10axv; mverd_mveds4; mdist_zds4}
if (k==17) {desdim_6; n_40; p_7; x_x6ax1; mverd_mveds5; mdist_zds5}
if (k==18) {desdim_6; n_40; p_7; x_x6ax1; mverd_mveds5; mdist_zds5}
if (k==19) {desdim_6; n_40; p_7; x_x6axv; mverd_mveds6; mdist_zds6}
if (k==20) {desdim_6; n_40; p_7; x_x6axv; mverd_mveds6; mdist_zds6}
if (k==21) {desdim_6; n_40; p_7; x_x6axv; mverd_mveds6; mdist_zds6}
if (k==22) {desdim_6; n_40; p_7; x_x6axv; mverd_mveds6; mdist_zds6}
if (k==23) {desdim_6; n_40; p_7; x_x6axv; mverd_mveds6; mdist_zds6}
if (k==24) {desdim_6; n_40; p_7; x_x6axv; mverd_mveds6; mdist_zds6}

#  Generate the vector of orthogonal coefficients

if (desdim==2) orthmag_7
if (desdim==6) orthmag_4
if (desdim==10) orthmag_3


alpha_orthmag + vector("numeric",length=(p-1))



#  Initialize the arrays that hold the coefficient estimates

mset_matrix(nrow=nrun,ncol=11)
lscf_matrix(nrow=nrun,ncol=p)
mcf_matrix(nrow=nrun,ncol=p)
rbmcf_matrix(nrow=nrun,ncol=p)
ltscf_matrix(nrow=nrun,ncol=p)
scf_matrix(nrow=nrun,ncol=p)
mmcf_matrix(nrow=nrun,ncol=p)
bichcf_matrix(nrow=nrun,ncol=p)
bimjcf_matrix(nrow=nrun,ncol=p)
bijs2cf_matrix(nrow=nrun,ncol=p)
bijs3cf_matrix(nrow=nrun,ncol=p)
bijs5cf_matrix(nrow=nrun,ncol=p)




for (i in 1:nrun) {


#
#  Create the error vector, dependent on the error distribution and
#  noise level
#

#
#  Determine the value of the residuals
#

resd_c(rnorm(n,0,1))



#
#  Modify the residual(s) to generate outlier(s)
#

#  Dataset 1 or 5 residuals --------------------------------------------------

if(k==1 || k==5) {
#  2 variable dataset - 10% (2) Interior outliers
#  finding the observation numbers for the outlier positions
#  must randomly select without replacement
    posset_NULL
    while (length(posset)<2)  {
    pos_floor(runif(1)*12)+1
    if (length(posset)==0) posset_pos else {
      cpre_compare(rep(pos,length(posset)),posset)
      if (any(cpre==0)) next
```

```
   posset_c(posset,pos)
   }
  }

#  assign a residual magnitude to the outlying observation

ressiz_c(8,10)
for (j in 1:2) {
 resd[posset[j]]_sign(resd[posset[j]])*ressiz[j]
 }
}
#  -----------------------------------------------------------------------


#  Dataset 2 or 6 residuals ---------------------------------------------------

if(k==2 || k==6) {
#  10 variable dataset - 10% (8) Interior outliers
#  finding the observation numbers for the outlier positions
#  must randomly select without replacement
  posset_NULL
  while (length(posset)<8)  {
  pos_floor(runif(1)*64)+1
  if (length(posset)==0) posset_pos else {
   cpre_compare(rep(pos,length(posset)),posset)
   if (any(cpre==0)) next
   posset_c(posset,pos)
   }
  }

#  assign a residual magnitude to the outlying observation

ressiz_c(12,14,16,18,12,14,16,18)
for (j in 1:8) {
 resd[posset[j]]_sign(resd[posset[j]])*ressiz[j]
 }
}
#  -----------------------------------------------------------------------


#  Dataset 3 or 7 residuals ---------------------------------------------------

if(k==3 || k==7) {
#  2 variable dataset - 20% (3) Interior outliers
#  finding the observation numbers for the outlier positions
#  must randomly select without replacement
  posset_NULL
  while (length(posset)<3)  {
  pos_floor(runif(1)*12)+1
  if (length(posset)==0) posset_pos else {
   cpre_compare(rep(pos,length(posset)),posset)
   if (any(cpre==0)) next
   posset_c(posset,pos)
   }
  }

#  assign a residual magnitude to the outlying observation

ressiz_c(6,8,10)
for (j in 1:3) {
 resd[posset[j]]_sign(resd[posset[j]])*ressiz[j]
 }
}
#  -----------------------------------------------------------------------


#  Dataset 4 or 8 residuals ---------------------------------------------------

if(k==4 || k==8) {
```

```
#  10 variable dataset - 20% (16) Interior outliers
#  finding the observation numbers for the outlier positions
#  must randomly select without replacement
    posset_NULL
    while (length(posset)<16)  {
    pos_floor(runif(1)*64)+1
    if (length(posset)==0) posset_pos else {
      cpre_compare(rep(pos,length(posset)),posset)
      if (any(cpre==0)) next
      posset_c(posset,pos)
      }
    }

#  assign a residual magnitude to the outlying observation

ressiz_c(12,14,16,18,12,14,16,18,12,14,16,18,12,14,16,18)
for (j in 1:16) {
  resd[posset[j]]_sign(resd[posset[j]])*ressiz[j]
  }
}
#  -----------------------------------------------------------------------


#  Dataset 9 residuals ---------------------------------------------------

if(k==9) {
#  2 variable dataset - 10% (2) Exterior outliers
#  finding the observation numbers for the outlier positions
#  must randomly select without replacement
    posset_NULL
    while (length(posset)<2)  {
    pos_floor(runif(1)*4)+13
    if (length(posset)==0) posset_pos else {
      cpre_compare(rep(pos,length(posset)),posset)
      if (any(cpre==0)) next
      posset_c(posset,pos)
      }
    }

#  assign a residual magnitude to the outlying observation

ressiz_c(8,10)
for (j in 1:2) {
  resd[posset[j]]_sign(resd[posset[j]])*ressiz[j]
  }
}
#  -----------------------------------------------------------------------


#  Dataset 10 residuals --------------------------------------------------

if(k==10) {
#  10 variable dataset - 10% (8) Exterior outliers
#  finding the observation numbers for the outlier positions
#  must randomly select without replacement
    posset_NULL
    while (length(posset)<8)  {
    pos_floor(runif(1)*16)+65
    if (length(posset)==0) posset_pos else {
      cpre_compare(rep(pos,length(posset)),posset)
      if (any(cpre==0)) next
      posset_c(posset,pos)
      }
    }

#  assign a residual magnitude to the outlying observation

ressiz_c(12,14,16,18,12,14,16,18)
```

```
for (j in 1:8) {
  resd[posset[j]]_sign(resd[posset[j]])*ressiz[j]
  }
}
# --------------------------------------------------------------------


# Dataset 11 residuals -------------------------------------------------

if(k==11) {
# 2 variable dataset - 20% (3) Exterior outliers
# finding the observation numbers for the outlier positions
# must randomly select without replacement
    posset_NULL
    while (length(posset)<3)  {
    pos_floor(runif(1)*4)+13
    if (length(posset)==0) posset_pos else {
      cpre_compare(rep(pos,length(posset)),posset)
      if (any(cpre==0)) next
      posset_c(posset,pos)
      }
    }

# assign a residual magnitude to the outlying observation

ressiz_c(6,8,10)
for (j in 1:3) {
  resd[posset[j]]_sign(resd[posset[j]])*ressiz[j]
  }
}
# --------------------------------------------------------------------


# Dataset 12 residuals -------------------------------------------------

if(k==12) {

# 10 variable dataset - 20% (16) Exterior outliers
# assign a residual magnitude to each Exterior outlying observation

ressiz_c(12,14,16,18,12,14,16,18,12,14,16,18,12,14,16,18)
for (j in 65:80) {
  resd[j]_sign(resd[j])*ressiz[j-64]
  }
}

# --------------------------------------------------------------------

# Dataset 13 residuals -------------------------------------------------

if(k==13) {
# 2 variable dataset - 10% (2) outliers (1 interior and 1 exterior)
# finding the observation numbers for the outlier positions
# must randomly select without replacement

# Assign the interior outlier
    posset_NULL
    while (length(posset)<1)  {
    pos_floor(runif(1)*12)+1
    if (length(posset)==0) posset_pos else {
      cpre_compare(rep(pos,length(posset)),posset)
      if (any(cpre==0)) next
      posset_c(posset,pos)
      }
    }

# assign a residual magnitude to the outlying observation
```

```
ressiz_c(8)
for (j in 1:1) {
  resd[posset[j]]_sign(resd[posset[j]])*ressiz[j]
  }

#  Assign the exterior outlier
    posset_NULL
    while (length(posset)<1)  {
    pos_floor(runif(1)*4)+13
    if (length(posset)==0) posset_pos else {
      cpre_compare(rep(pos,length(posset)),posset)
      if (any(cpre==0)) next
      posset_c(posset,pos)
      }
    }

#  assign a residual magnitude to the outlying observation

ressiz_c(10)
for (j in 1:1) {
  resd[posset[j]]_sign(resd[posset[j]])*ressiz[j]
  }
}
#  -------------------------------------------------------------------

#  Dataset 14 residuals ------------------------------------------------

if(k==14) {
#  10 variable dataset - 10% (8) outliers (4 interior and 4 exterior)
#  finding the observation numbers for the outlier positions
#  must randomly select without replacement

#  Assign the interior outlier
    posset_NULL
    while (length(posset)<4)  {
    pos_floor(runif(1)*64)+1
    if (length(posset)==0) posset_pos else {
      cpre_compare(rep(pos,length(posset)),posset)
      if (any(cpre==0)) next
      posset_c(posset,pos)
      }
    }

#  assign a residual magnitude to the outlying observation

ressiz_c(12,14,16,18)
for (j in 1:4) {
  resd[posset[j]]_sign(resd[posset[j]])*ressiz[j]
  }

#  Assign the exterior outlier
    posset_NULL
    while (length(posset)<4)  {
    pos_floor(runif(1)*16)+65
    if (length(posset)==0) posset_pos else {
      cpre_compare(rep(pos,length(posset)),posset)
      if (any(cpre==0)) next
      posset_c(posset,pos)
      }
    }

#  assign a residual magnitude to the outlying observation

ressiz_c(12,14,16,18)
for (j in 1:4) {
  resd[posset[j]]_sign(resd[posset[j]])*ressiz[j]
  }
}
```

```
#  ------------------------------------------------------------------------

#  Dataset 15 residuals ---------------------------------------------------

if(k==15) {
#  2 variable dataset - 20% (3) outliers (1 interior and 2 exterior)
#  finding the observation numbers for the outlier positions
#  must randomly select without replacement

#  Assign the interior outlier
    posset_NULL
    while (length(posset)<1)  {
    pos_floor(runif(1)*12)+1
    if (length(posset)==0) posset_pos else {
      cpre_compare(rep(pos,length(posset)),posset)
      if (any(cpre==0)) next
      posset_c(posset,pos)
      }
    }

#  assign a residual magnitude to the outlying observation

ressiz_c(6)
for (j in 1:1) {
  resd[posset[j]]_sign(resd[posset[j]])*ressiz[j]
  }

#  Assign the exterior outlier
    posset_NULL
    while (length(posset)<2)  {
    pos_floor(runif(1)*4)+13
    if (length(posset)==0) posset_pos else {
      cpre_compare(rep(pos,length(posset)),posset)
      if (any(cpre==0)) next
      posset_c(posset,pos)
      }
    }

#  assign a residual magnitude to the outlying observation

ressiz_c(8,10)
for (j in 1:2) {
  resd[posset[j]]_sign(resd[posset[j]])*ressiz[j]
  }
}
#  ------------------------------------------------------------------------

#  Dataset 16 residuals ---------------------------------------------------

if(k==16) {
#  10 variable dataset - 20% (16) outliers (8 interior and 8 exterior)
#  finding the observation numbers for the outlier positions
#  must randomly select without replacement

#  Assign the interior outlier
    posset_NULL
    while (length(posset)<8)  {
    pos_floor(runif(1)*64)+1
    if (length(posset)==0) posset_pos else {
      cpre_compare(rep(pos,length(posset)),posset)
      if (any(cpre==0)) next
      posset_c(posset,pos)
      }
    }

#  assign a residual magnitude to the outlying observation
```

```
ressiz_c(12,14,16,18,12,14,16,18)
for (j in 1:8) {
  resd[posset[j]]_sign(resd[posset[j]])*ressiz[j]
  }

#  Assign the exterior outlier
    posset_NULL
    while (length(posset)<8)  {
    pos_floor(runif(1)*16)+65
    if (length(posset)==0) posset_pos else {
      cpre_compare(rep(pos,length(posset)),posset)
      if (any(cpre==0)) next
      posset_c(posset,pos)
      }
    }

#  assign a residual magnitude to the outlying observation

ressiz_c(12,14,16,18,12,14,16,18)
for (j in 1:8) {
  resd[posset[j]]_sign(resd[posset[j]])*ressiz[j]
  }
}
#  ----------------------------------------------------------------------

#  Dataset 17 or 19 residuals --------------------------------------------------

if(k==17 || k==19) {
#  6 variable dataset - 10% (4) Interior outliers
#  finding the observation numbers for the outlier positions
#  must randomly select without replacement
    posset_NULL
    while (length(posset)<4)  {
    pos_floor(runif(1)*32)+1
    if (length(posset)==0) posset_pos else {
      cpre_compare(rep(pos,length(posset)),posset)
      if (any(cpre==0)) next
      posset_c(posset,pos)
      }
    }

#  assign a residual magnitude to the outlying observation

ressiz_c(8,10,12,14)
for (j in 1:4) {
  resd[posset[j]]_sign(resd[posset[j]])*ressiz[j]
  }
}
#  ----------------------------------------------------------------------


#  Dataset 18 or 20 residuals --------------------------------------------------

if(k==18 || k==20) {
#  6 variable dataset - 20% (8) Interior outliers
#  finding the observation numbers for the outlier positions
#  must randomly select without replacement
    posset_NULL
    while (length(posset)<8)  {
    pos_floor(runif(1)*32)+1
    if (length(posset)==0) posset_pos else {
      cpre_compare(rep(pos,length(posset)),posset)
      if (any(cpre==0)) next
      posset_c(posset,pos)
      }
    }

#  assign a residual magnitude to the outlying observation
```

```
ressiz_c(8,10,12,14,8,10,12,14)
for (j in 1:8) {
  resd[posset[j]]_sign(resd[posset[j]])*ressiz[j]
  }
}
#  ----------------------------------------------------------------------

#  Dataset 21 residuals ---------------------------------------------------

if(k==21) {
#  6 variable dataset - 10% (4) Exterior outliers
#  finding the observation numbers for the outlier positions
#  must randomly select without replacement
    posset_NULL
    while (length(posset)<4)  {
    pos_floor(runif(1)*8)+33
    if (length(posset)==0) posset_pos else {
      cpre_compare(rep(pos,length(posset)),posset)
      if (any(cpre==0)) next
      posset_c(posset,pos)
      }
    }

#  assign a residual magnitude to the outlying observation

ressiz_c(8,10,12,14)
for (j in 1:4) {
  resd[posset[j]]_sign(resd[posset[j]])*ressiz[j]
  }
}
#  ----------------------------------------------------------------------



#  Dataset 22 residuals ---------------------------------------------------

if(k==22) {

#  6 variable dataset - 20% (8) Exterior outliers
#  assign a residual magnitude to each Exterior outlying observation

ressiz_c(8,10,12,14,8,10,12,14)
for (j in 33:40) {
  resd[j]_sign(resd[j])*ressiz[j-32]
  }
}

#  ----------------------------------------------------------------------


#  Dataset 23 residuals ---------------------------------------------------

if(k==23) {
#  6 variable dataset - 10% (4) outliers (2 interior and 2 exterior)
#  finding the observation numbers for the outlier positions
#  must randomly select without replacement

#  Assign the interior outlier
    posset_NULL
    while (length(posset)<2)  {
    pos_floor(runif(1)*32)+1
    if (length(posset)==0) posset_pos else {
      cpre_compare(rep(pos,length(posset)),posset)
      if (any(cpre==0)) next
      posset_c(posset,pos)
      }
    }
```

```
#  assign a residual magnitude to the outlying observation

ressiz_c(8,12)
for (j in 1:2) {
  resd[posset[j]]_sign(resd[posset[j]])*ressiz[j]
  }

#  Assign the exterior outlier
    posset_NULL
    while (length(posset)<2)  {
    pos_floor(runif(1)*8)+33
    if (length(posset)==0) posset_pos else {
      cpre_compare(rep(pos,length(posset)),posset)
      if (any(cpre==0)) next
      posset_c(posset,pos)
      }
    }

#  assign a residual magnitude to the outlying observation

ressiz_c(10,14)
for (j in 1:2) {
  resd[posset[j]]_sign(resd[posset[j]])*ressiz[j]
  }
}
#  ---------------------------------------------------------------------

#  Dataset 24 residuals ------------------------------------------------

if(k==24) {
#  6 variable dataset - 20% (8) outliers (4 interior and 4 exterior)
#  finding the observation numbers for the outlier positions
#  must randomly select without replacement

#  Assign the interior outlier
    posset_NULL
    while (length(posset)<4)  {
    pos_floor(runif(1)*32)+1
    if (length(posset)==0) posset_pos else {
      cpre_compare(rep(pos,length(posset)),posset)
      if (any(cpre==0)) next
      posset_c(posset,pos)
      }
    }

#  assign a residual magnitude to the outlying observation

ressiz_c(8,10,12,14)
for (j in 1:4) {
  resd[posset[j]]_sign(resd[posset[j]])*ressiz[j]
  }

#  Assign the exterior outlier
    posset_NULL
    while (length(posset)<4)  {
    pos_floor(runif(1)*8)+33
    if (length(posset)==0) posset_pos else {
      cpre_compare(rep(pos,length(posset)),posset)
      if (any(cpre==0)) next
      posset_c(posset,pos)
      }
    }

#  assign a residual magnitude to the outlying observation

ressiz_c(8,10,12,14)
for (j in 1:4) {
  resd[posset[j]]_sign(resd[posset[j]])*ressiz[j]
  }
}
#  ------------------------------------------------------------------



#
#  Generate the vector of observations
#

y_c(x %*% alpha + resd)

#
#  Compute INITIAL estimates for bounded influence techniques
#

ltscf[i,]_ltsreg(x,y)$coef
sout_sest(x,y)  #  Need the entire object so the scale estimate can be passed
scf[i,]_sout$coef
xwi_cbind(1,x)
dfrpar(xwi,"Kra-Wel")
rbmcf[i,]_rbmost(xwi,y,cc=1.5)$theta[1:p]


#
#  Compute FINAL estimates for bounded influence techniques
#

lscf[i,]_lsfit(x, y)$coef
mcf[i,]_rreg(x, y, method=wt.bisquare)$coef
mmcf[i,]_mmest(x, y)$coef
bichcf[i,]_ bich (x, y, init=ltscf[i,],  mverd=mverd)$coef
bimjcf[i,]_ bimj (x, y, init=rbmcf[i,],  mdist=mdist)$coef
bijs2cf[i,]_bijs2(x, y, init=sout,  mdist=mdist)$coef
bijs3cf[i,]_bijs3(x, y, init=sout,  mdist=mdist)$coef
bijs5cf[i,]_bijs5(x, y, init=sout,  mdist=mdist)$coef


talpha_c(0,alpha)

#  Calculate the mean square inefficiency ratios

msels_t(lscf[i,]-talpha)%*%(lscf[i,]-talpha)
msem_t(mcf[i,]-talpha)%*%(mcf[i,]-talpha)
mserbm_t(rbmcf[i,]-talpha)%*%(rbmcf[i,]-talpha)
mselts_t(ltscf[i,]-talpha)%*%(ltscf[i,]-talpha)
mses_t(scf[i,]-talpha)%*%(scf[i,]-talpha)
msemm_t(mmcf[i,]-talpha)%*%(mmcf[i,]-talpha)
msebich_t(bichcf[i,]-talpha)%*%(bichcf[i,]-talpha)
msebimj_t(bimjcf[i,]-talpha)%*%(bimjcf[i,]-talpha)
msebijs2_t(bijs2cf[i,]-talpha)%*%(bijs2cf[i,]-talpha)
msebijs3_t(bijs3cf[i,]-talpha)%*%(bijs3cf[i,]-talpha)
msebijs5_t(bijs5cf[i,]-talpha)%*%(bijs5cf[i,]-talpha)

mset[i,]_c(msels,msem,mserbm,mselts,mses,msemm,msebich,msebimj,msebijs2,msebijs3,msebijs5)

#  Close the loop that performs tests of multiple samples for fixed settings
}

#  Calculate the mean mse for each technique

amsels_mean(mset[,1])
amsem_mean(mset[,2])
amserbm_mean(mset[,3])
amselts_mean(mset[,4])
amses_mean(mset[,5])
amsemm_mean(mset[,6])
```

```
amsebich_mean(mset[,7])
amsebimj_mean(mset[,8])
amsebijs2_mean(mset[,9])
amsebijs3_mean(mset[,10])
amsebijs5_mean(mset[,11])

amse_c(amsels,amsem,amserbm,amselts,amses,amsemm,amsebich,amsebimj,amsebijs2,amsebijs3,amsebijs5)
write(t(amse),file="p5tamse.out",ncol=length(amse),append=TRUE)


# Write the contents of mse estimation to a file by rows

if (k==1) write(t(mset),file="p5tds1.out",ncol=ncol(mset))
if (k==2) write(t(mset),file="p5tds2.out",ncol=ncol(mset))
if (k==3) write(t(mset),file="p5tds3.out",ncol=ncol(mset))
if (k==4) write(t(mset),file="p5tds4.out",ncol=ncol(mset))
if (k==5) write(t(mset),file="p5tds5.out",ncol=ncol(mset))
if (k==6) write(t(mset),file="p5tds6.out",ncol=ncol(mset))
if (k==7) write(t(mset),file="p5tds7.out",ncol=ncol(mset))
if (k==8) write(t(mset),file="p5tds8.out",ncol=ncol(mset))
if (k==9) write(t(mset),file="p5tds9.out",ncol=ncol(mset))
if (k==10) write(t(mset),file="p5tds10.out",ncol=ncol(mset))
if (k==11) write(t(mset),file="p5tds11.out",ncol=ncol(mset))
if (k==12) write(t(mset),file="p5tds12.out",ncol=ncol(mset))
if (k==13) write(t(mset),file="p5tds13.out",ncol=ncol(mset))
if (k==14) write(t(mset),file="p5tds14.out",ncol=ncol(mset))
if (k==15) write(t(mset),file="p5tds15.out",ncol=ncol(mset))
if (k==16) write(t(mset),file="p5tds16.out",ncol=ncol(mset))
if (k==17) write(t(mset),file="p5tds17.out",ncol=ncol(mset))
if (k==18) write(t(mset),file="p5tds18.out",ncol=ncol(mset))
if (k==19) write(t(mset),file="p5tds19.out",ncol=ncol(mset))
if (k==20) write(t(mset),file="p5tds20.out",ncol=ncol(mset))
if (k==21) write(t(mset),file="p5tds21.out",ncol=ncol(mset))
if (k==22) write(t(mset),file="p5tds22.out",ncol=ncol(mset))
if (k==23) write(t(mset),file="p5tds23.out",ncol=ncol(mset))
if (k==24) write(t(mset),file="p5tds24.out",ncol=ncol(mset))

# Close the loop for dataset tests
}
```